THE SCIENCE OF MARKING

By TERRY THOMAS, M.A., Ph.D., B.Sc., LL.B.

HEAD MASTER OF LEEDS GRAMMAR SCHOOL; LATE HEAD OF THE MILITARY AND ENGINEERING SIDE, HAILEYBURY COLLEGE; FORMERLY SCHOLAR OF ST. JOHN'S COLLEGE, CAMBRIDGE

This book is an attempt to deal with marks in a scientific manner.

Methods of marking are discussed and graphical methods are used to illustrate results.

An introduction is given to modern statistical methods of dealing with marks and numerical returns.

The methods are fully and simply explained and examples of their application are given.

The book should be useful to all engaged in educational work and to Training Colleges.

The methods here applied to school marks may be used equally well by business men and administrators in arranging their own returns and statistics.

4s. 6d. net

LONDON: JOHN MURRAY

THE SCIENCE OF MARKING

THE SCIENCE OF MARKING

BY

TERRY THOMAS, M.A., Ph.D., B.Sc., LL.B.

HEAD MASTER OF LEEDS GRAMMAR SCHOOL;
LATE HEAD OF THE MILITARY AND ENGINEERING SIDE, HAILEYBURY COLLEGE;
FORMERLY SCHOLAR OF ST. JOHN'S COLLEGE, CAMBRIDGE

LONDON
JOHN MURRAY, ALBEMARLE STREET, W.

First Edition . . . 1930

PREFACE

MARKS and examinations enter so largely into the life of the average student that it is fitting that the whole matter should be looked at from a scientific point of view. An attempt is made in this book to deal with some of the problems and difficulties which arise.

It is not suggested that all masters should work out their marks with such elaboration, or that marks should be given undue prominence, but if an order of merit has to be worked out, it should be done scientifically or it will be untrustworthy. This is especially true in competitive examinations and in all cases where rewards such as scholarships, promotions, etc., depend on the order of the candidates. Further, if any reliable deductions are to be made from Intelligence Tests and Standard Tests, it is of the utmost importance that the marking and the results should be dealt with in a logical and systematic way.

It is hoped that administrators will find that some of the methods outlined in this book can be used effectively in arranging their own returns and statistics.

T. T.,
Leeds.

June, 1930.

CONTENTS

CHAP.		PAGE
I	EXAMINATIONS AND MARKS	1
	Examinations	1
	Marks	4
	The Main Use of Marks	8
	Marking	9
	Methods of Marking	10
	Compounding Marks	11
	Form Marks	13
	Scaling	17
II	GRAPHICAL METHODS	23
	Histograms and Frequency Curves	23
	Percentile Curves	27
III	AVERAGES AND NORMS	30
	The Average	30
	The Mode	31
	The Median	32
	The Upper Quartile	32
	The Lower Quartile	32
	Norms	33
	The Intelligence Quotient	34
	Standard Tests	36
IV	CORRELATIONS	39
	Correlation	39
	The Method of Ranks	42
	The Footrule Method	47
	The Product Moment Method	49
	Probable Error	50
	Semi-interquartile Range	51
	Coefficient of Variability	52
	Probable Error of Correlation Coefficient	53
	Summary of P.E. for Correlation Coefficients	55

CHAP.		PAGE
	Correlation of Intelligence Test and General Knowledge Test	57
	Reliability of a Test	58
	Rough Methods of Finding Relationship	61
	(*a*) Pearson's Method	61
	(*b*) Shepherd's Method	62
	(*c*) Coefficient of Association	63
	(*d*) Coefficient of Colligation	63
	Comparison of Results	66
V	MANAGEMENT OF SCHOOL MARKS	67
	Form Returns	67
	Analysis into Score Groups	73
	Percentile Curves for each Age Group	80
	Average Scores	81
	Comparison of Average with Median	82
	Percentile and Histogram Curves for an Intelligence Test	82
	Percentile Curves for each Age Group	85
	Comparison of Forms	86
	Latin Grammar Paper	88
	Accuracy Paper	96
	French Grammar Paper	99
	Percentile Record Card	103
VI	SOME FURTHER APPLICATIONS	104
	Allowance for Age	104
	School Estimates and Examination Results	110
	Correlation as a Test	110
	Comparing the Standards	114
	APPENDIX	117
	The Normal Curve	117
	The Ideal Mark Scale	119
	INDEX	124

CHAPTER I

EXAMINATIONS AND MARKS

EXAMINATIONS

THOUGH examinations are subject to much abuse and criticism, no one has yet suggested a reasonable substitute. Not only is entrance to the Civil Service and admission to the University governed by examinations, but the system of examinations is extending to all the professions. Whatever profession a boy chooses, whether it be that of a Banker, Insurance Official, Architect or Surveyor, he is now faced with passing professional examinations. Again, some of the most valuable prizes, such as scholarships, are given on the results of competitive examinations, and the award of these is often determined solely by the marks obtained. Hence the marking system is of fundamental importance, and it does not always get the amount of thought and consideration which its importance deserves.

There is also a tendency to judge schools and teachers by examination results, and this may lead to very grave injustice. It is obviously much easier to prepare a class of good candidates for an examination than it is to prepare a class of inferior ones. It might be more creditable for a teacher to get an average mark of 40 per cent with an inferior class than for another to get an average mark of 60 per cent with a superior class. Yet the latter teacher

would be counted as more successful in the opinion of the uninformed public. Thus, though the examination results are useful for comparative purposes, their use should not be extended without consideration of the other factors.

It is impossible to compare schools or teachers by the examination results they produce without also comparing the intellectual quality of the material with which they have to deal. It is a fallacy, moreover, to assume that a 33 per cent standard in a given subject one year will be the same as a 33 per cent standard in the same subject another year. It is obvious that the marks depend on the papers and on the examiners.

Suppose that a local authority decided to give scholarships to all candidates who reached a mark of 60 per cent. In one year the papers might be easy and the examiners generous, and so a large number of candidates would get scholarships, whereas in another year, if the papers were hard or badly set, the number of candidates who reached the 60 per cent mark might be reduced and therefore fewer scholarships would be awarded.

It is particularly dangerous to compare the percentage marks of one year with another unless the papers are of the same difficulty exactly and the marking scale also the same. But who is to say whether the papers are of the same difficulty or not? It is practically impossible to make sure of that. Even experienced examiners and teachers will differ amongst themselves. In large examinations, such as the School Certificate Examination, marks are sometimes adjusted so as to get the same number of candidates above a certain percentage. An illustration will make this clear.

Let us suppose it is usual to give a credit to approximately one half of the candidates and that there are 14,000 candidates in a subject in 1927, out of which 7,000 candidates reach 45 per cent. Thus one half of the candidates have reached the credit mark of 45 per cent. If, on the other hand, in 1928, out of 15,000 candidates only 6,000 reached 45 per cent, it is obvious that the examination is harder or the marking has been more severe. In 1928 it might be found that one has to go down to a mark of 39 per cent in order to draw the line, so that half of the candidates will have passed. Thus a 45 per cent mark in 1927 corresponds to a 39 per cent mark in 1928. The 1928 marks can then be adjusted so that the original mark of 39 per cent is brought up to 45 per cent.

The simplest way of doing this is to multiply all the candidates' marks by $\frac{45}{39}$, or they may be scaled by means of a graph. This is tantamount to increasing the value of the unit mark which the examiners have been giving. It may be argued that this should not be done and that in 1928 the candidates were weaker. But when—as in the School Certificate Examination—candidates are drawn from so many schools, it is unlikely that there should be such a general slump in the work, and the more reasonable conclusion is to suppose that the marks have been harder to get. The important function of the examiners is, first, to determine the order and afterwards to settle where the line should be drawn. This is by no means an easy task because the competition round about the line will be very severe, and to help them in finally drawing the line examining bodies should pay attention to school records.

In the Northern Universities Joint Board Examinations school records are taken into account, especially in the case of candidates who are just below the borderline. Each school submits an estimate of the marks a candidate is expected to get, and though these estimates will vary in reliability from school to school, they are useful in determining the relative order of merit, quite apart from the actual mark allotted by the school. Thus it is the boy's position relative to his fellows which is of importance in the school record, for the examining board can then estimate whether the slight failure of a candidate is due to the fact that he has not done himself justice. If he has just failed and yet in the school record appears high in the list, the examining board can give the candidate the benefit of the doubt and thus the school record is taken into account.

It should be borne in mind, however, that the examining board can check the reliability of the school estimate by means of the rank correlation (see p. 42). If this is high, the board would naturally be more inclined to take the school record into account in dealing with the borderline cases. If the school order of merit does not correspond at all with the results of the examination, then the examining board will have every reason to be cautious in making allowances for school records.

MARKS

School marks form the common ground on which parents, pupils and pedagogues meet. In most schools subject or form marks are issued during the School Year, sometimes with tantalizing frequency.

It would take too long to describe with the necessary vigour the views of some schoolmasters and also

some schoolboys about the mark system, but there is no doubt that the mark system is useful, because we have to handle a number which is easily appreciated by the minds of all the patient "Trinity" outlined above, instead of such vague terms as "fair" and "very fair", and so on. The significance of the terms "fair" and "moderate", which occur so often on School Reports, depends so much on the standard of the person making the report. Such expressions are vague and indefinite and are liable to misconstruction. The following story illustrates this. A parent was buoyed up with great hope at half term on being informed that his boy was "trying", only to be cast down to the depths by the End-of-Term Report which described his young hopeful as "very trying".

Hence all who wish to have some reliable footrule by which to measure the progress or the standard of a pupil have to resort to marks, and for this reason marks have been in use in the schools for generations. In some cases a standard scale has been expressed in terms of Alphas, Betas and Gammas, and some modern schools substitute for the vicious mark system beautiful systems of A's, B's and C's. Many of these new mark systems soon degenerate into a sequence of—

A +; A; A −; B +; B; B −; C +; C; C −.

This practically amounts to a nine-point scale, and sometimes, from a perverted sense of humour, D is added at the bottom of the scale to give pathos and significance, whilst at the same time it practically converts the scale into a range of marks from 10 to 1.

This scale of A's and B's has all the disadvantages

of any mark scale without the advantages. For the great advantage of a mark scale is that marks for separate tests or separate subjects can be added together and an order produced, whereas it is beyond the wit of man to decide whether (A +) + (B) + (C) is better than (A −) + (B +) + (C +).

So long as orders have to be produced marks seem essential, and the importance of arranging a form or a number of candidates in order of merit or in ranks is increasing every day, owing to the advent of Intelligence Tests and Standardized Tests. It is possible, though inconvenient, to do without marks in a single subject, but it seems impossible to deal with aggregates extending over a number of pupils taking a number of subjects, without a numerical system.

It is accepted at the moment that a boy with a mark of 61 is better than a boy with a mark of 60, though (as is shown later) there may be some reason to doubt this. After all, markers are not infallible and, like the rest of us, are subject to the chances and changes of this wicked world.

It does not follow that a mark gained for a piece of work in one part of a paper is equal in value to a mark gained for another piece of work in another part of the paper. This difficulty even occurs in one and the same question. Suppose, for example, in an answer to a history question, there are ten relevant points which might be brought in and that one boy brings in five of them and another brings in the other five. Are we to mark these answers at an equal value? It is very difficult to say without some knowledge of the relative importance of these points, and we can only leave the matter to the judgment of the examiners. For examination and

school purposes, it is generally accepted that a higher mark for a piece of work signifies that this piece of work is better than another piece of work which gets a lower mark. The whole difficulty is that pieces of work may not be strictly comparable, and so we have to do the best we can in "rating" their value. Thus marks in many cases become the expression of the personal judgment of the marker.

Perhaps the day will come when we shall obtain some objective marking instrument for assessing marks, in the same way that a thermometer registers temperature. For all practical purposes, a standard thermometer has a uniform scale so that the unit of temperature is the same at all parts of the scale. Unfortunately, we have no such thing in marks, though perhaps some day we shall get a suitable scientific mark system with a unit mark, in the same way as we have a standard temperature scale.

An exact scientific mark system would make a step of one mark on the scale equal on all points of the scale. Such a mark would be almost a work unit, and the work required by a candidate to raise his position one unit on the scale would be the same at all parts of the scale. This would be analogous to the fact that, in mechanics, the work required to raise a weight one unit is the same whether the weight be on the floor or 20 centimetres above the floor, or 60 centimetres above the floor. Unfortunately, with marks this is not so. Marks may be easy to get on certain questions and hard to get on others, so that a pupil may not require so much work or so much ability to jump from 20 to 21 as another pupil does to jump from 60 to 61. Where a paper contains a number of questions, it is obvious that the questions may not be of equal value. In some cases no attempt

is made to weight the questions according to their value, whilst in other cases the weighting is only understood by him who stands and weights.

THE MAIN USE OF MARKS

(*a*) *To spread the candidates so that an order of merit can be obtained.*

This is the most obvious use of marks, and thereby we are enabled to put the candidates in order from the first down to the last. It is necessary to have these marks, or the order of merit, if such a list is to be compared with any other results obtained by the same candidates. These methods of comparison will be discussed later under the heading of Correlations. The correlation coefficient can be calculated rigidly from the marks by the ordinary statistical methods, but if the marks are not used, it is possible, provided the orders of merit are known, to calculate the correlation by the method of ranks due to Professor Spearman.

(*b*) *To find how many candidates reach a certain level.*

Marks are continually used for this purpose. For instance, in a School Certificate Examination, a mark of 45 per cent may be required to obtain a credit and a mark of 33 per cent to obtain a pass. So if the candidates are arranged in order and a line is drawn under the one with 45 per cent, all the candidates above this line have reached a credit standard. If lower down a line is drawn under the candidate with 33 per cent, all the candidates between these two lines have reached the pass standard and those below the bottom line have failed.

To prevent too great a variation from year to year in the standard required for credits and passes, some

examining bodies standardize the marks. This is done by giving approximately the same percentage of credits and the same percentage of passes year by year. This is probably more satisfactory than allowing the examiners to set their own absolute standard, since the variations might otherwise be serious.

There are, of course, difficulties about standardization. If a large number of badly prepared candidates are put in, standardization means that it is easier in that year to get a credit or a pass. On the other hand, as the efficiency of the schools increases and if the candidates are selected, then it is getting harder year by year to reach the required standard. Standardization means then a gradual raising of the standard as the standard of the examinees rises. This possibly gives an advantage to the large and good schools at the expense of candidates from the schools with fewer facilities.

Thus there are dangers in standardization, but it seems better, on the whole, to be at the mercy of standardization than to be completely in the hands of the examiners, who may have peculiar and variable standards.

MARKING

Marking is a particularly difficult business, especially if an examiner has to mark a large number of papers. It is usual to settle the marks to be allotted to each question before the papers are marked, though sometimes the examiners take a sample of the papers and correct them without giving marks, and by this initial sampling they are able to assess roughly the mark value of each question.

Even when the mark scale is settled, there must be

a tendency to vary the marks awarded because the marker may be more alert at one time than another. This is a great difficulty in marking. After two or three hours of marking papers, it is hardly likely that the examiner will be as fresh or as alert as when he started. On the other hand, he is by this time aware of the common errors. Whether this " awareness " balances his lack of freshness it is difficult to say, but it is exceedingly probable that his standard is not quite the same. This variation of standard is most likely to occur in marking literary questions or questions of the essay type. It is possible to split up the marks for a mathematical question such as one mark for this part and two for that part, and so on—so that there is not much departure from the fair scale, even with a tired examiner. But in the case of an essay or history answer, there must be variations. The answers vary so much amongst themselves, some boys putting in what others leave out and vice versa. If a number of essays were given to different examiners to place in order, it would be found that the results would vary a great deal.

Many instances can be given to show the great difficulty that must arise in marking papers.

METHODS OF MARKING

Again, methods of marking vary. Some examiners read through an answer and say, " That answer is only worth 8 out of 10." Others may take off a mark for every mistake made and then by subtracting from the maximum allot a mark to the candidate. Others again may give marks for answers that are right. This method is often used in mathematical questions, where a boy may not have solved the whole question but is given a certain number of marks

because he has gone through certain stages in the answer.

These various mark systems will depend on the subject—for the mark system which applies to mathematics or to Latin sentences may not lend itself to the marking of English Literature or History questions.

Obviously, then, not only does the personal feeling of the marker come in, but the system which he uses may influence the final order. By marking a given set of papers by each system the systems can be compared by correlating the results, and if the correlations are good this would indicate that each mark system produces a reasonable and just order amongst the candidates. The system which is probably liable to the greatest error is the general-impression method, though, of course, some men have great ability in this direction and are able to sum up almost at a glance the value of an answer. Enough has been said, however, to show that the personal equation of the examiner and the system of marking may have appreciable effects in the final results.

COMPOUNDING MARKS

Another great difficulty arises when marks for different subjects are added together. Should each subject carry equal weight ? Let us take the simple case of two papers—one in Arithmetic and one in English.

TABLE I

CASE I.—ENGLISH AND ARITHMETIC BOTH CARRYING 100 MARKS MAXIMUM

Name.	English.	Arithmetic.	Total.	Rank.
A . . .	72	47	119	4
B . . .	71	50	121	2
C . . .	68	52	120	3
D . . .	74	48	122	1

CASE II.—ENGLISH 100, ARITHMETIC 200

Name.	English.	Arithmetic.	Total.	Rank.
A . . .	72	94	166	4
B . . .	71	100	171	2
C . . .	68	104	172	1
D . . .	74	96	170	3

This means doubling the Arithmetic marks.

CASE III.—ENGLISH 200, ARITHMETIC 100

Name.	English.	Arithmetic.	Total.	Rank.
A . . .	144	47	191	3
B . . .	142	50	192	2
C . . .	136	52	188	4
D . . .	148	48	196	1

This means doubling the English marks.

If 100 marks are allotted as a maximum for each paper and the marks are added together, an order will be produced. If the examiner in English has been more generous than the examiner in Arithmetic, then he will influence the final order more than his colleague who has marked the Arithmetic. Also, it will be found in actual practice that if 200 marks be allotted for the Arithmetic and 100 for the English, then the final order produced will differ from that obtained when the subjects carried equal weight. Again, another order will be produced if 200 be allotted for English and 100 for Arithmetic. Which order is the correct one? This may not be of great importance in ordinary form work, or even in a qualifying examination, but it is of the utmost importance in a competitive examination. If eighty candidates are competing for six scholarships, the weight which is given to the Arithmetic and the English may settle which boys are going to get the fifth and sixth scholarships, and so the results are partly determined by the weight which is given to each subject.

It is possible to get some idea which subject is the better guide by correlating the Arithmetic results and the English results with the order obtained by considering the later school careers of the boys. An example will make this clear. The order of a number of boys in a Form in 1927 was compared with their order in the various papers of the Scholarship Examination in 1925. It seems a fair assumption to make that the order of general ability in 1927 is more likely to be reliable than that obtained in the Scholarship papers. It was found that the correlations of the results of the Scholarship Examination and the 1927 Form Order were as follows:

Arithmetic, 0·40. Essay, 0·50. General Paper, 0·57.

Therefore it would have been wiser—had one but known it—to have allowed the General Paper in 1925 to have carried greater weight, for the General Paper order would have been a better guide to the general ability of the boys. If it could be shown over a space of years that the General Paper had always been a better guide, then the case would be established for allowing the General Paper marks to count more. This, however, might be upset by a bad General Paper, and in one year it might be found that the Arithmetic Paper was a better guide. Since the main object of a Scholarship Examination is to select candidates of promise, the papers should be set carefully and of the same standard year by year, and then a consideration of the school career of the candidates would throw some light on the question as to which of the Scholarship papers is the better "pointer" for future success.

FORM MARKS

Enough has been said to show the difficulty in arriving at a true order amongst a number of can-

didates taking a number of subjects. One common case where this difficulty is always arising is in the question of Form Orders.

Suppose there are thirty boys in a Form taking nine subjects and gaining marks in each subject. How should those marks be dealt with so as to produce a final order? It is obvious that the "raw" marks cannot be added together. Consider the following Table:

TABLE II

RAW MARKS

(The maximum possible in that week is indicated in the bracket.)

A.

Name.	English. (60)	History. (140)	French. (100)	Latin. (150)	Maths. (120)	Sc. (40)	Total.	Rank.
A	42	98	71	68	71	31	381	2
B	48	120	28	42	84	25	347	3
C	21	40	43	41	98	27	270	9
D	35	70	82	91	88	30	396	1
E	30	60	57	50	65	18	280	7
F	51	75	32	30	47	40	275	8
G	50	91	48	40	49	32	310	5
H	20	35	32	28	75	25	215	10
I	46	82	30	20	91	17	286	6
J	42	110	41	45	84	19	341	4

B.

Maximum Possible Marks Proportional to Number of Lessons per Week.

Name.	English. (30)	History. (20)	French. (50)	Latin. (50)	Maths. (60)	Sc. (40)	Total.	Rank.
A	21	14	36	23	36	31	161	2
B	24	17	14	14	42	25	136	3
C	11	6	22	14	49	27	129	6
D	18	10	41	30	44	30	173	1
E	15	9	26	17	33	18	118	9
F	26	11	16	10	24	40	127	7
G	25	13	24	13	25	32	132	5
H	10	5	16	9	38	25	103	10
I	23	12	15	7	46	17	120	8
J	21	16	21	15	42	19	134	4

The marks gained in each subject by ten pupils are added together and the order found. It will be seen that History, though it occupies only a few lessons in the week, has exerted a preponderating influence on the order because the marks for that subject are high. Thus it pays a boy to concentrate upon his History rather than on Latin, because it is easier to get marks in History. The boy who is good at History has a pull.

Let us see what happens to the final order if 10 marks are allotted for each lesson in the week. Then History, with two lessons, would carry a maximum of 20; Latin, with five lessons, would carry a maximum of 50; Mathematics, with six lessons, would carry a maximum of 60, and so on. The marks of the respective subjects must now be rearranged so that there is a maximum of 20 for History, 50 for Latin and 60 for Mathematics. The marks are then added together and a new final order is produced. (See Table IIB.) It will be seen to differ from the order produced by adding the "raw" marks together. We have taken a case of only ten boys, but in a Form of thirty boys it will be found that many more changes in the ranking occur. Therefore it is important to settle how the marks are to be allotted amongst the various subjects in the School Time Table. We are faced with the same problem as we were in the competitive examination as to what weight each subject should carry. It seems reasonable in the School Time-Table to allot the marks on a basis of the number of lessons, for if the available school time is divided amongst the subjects in definite and settled proportions, there is much to be said for allotting the marks in the same proportion.

There are two further questions which arise. Should the " raw " marks out of the settled maximum be added together, or should the marks be scaled so that the top boy in each subject gets a maximum mark ? This will be dealt with in Scaling (see page 19, Table III). A further and more difficult question is the arrangement of marks where there are a number of Sets and a given Set may include boys from different Forms. Suppose, for simplicity, there are three Latin Sets in which boys have been arranged according to their ability at Latin. Should the bottom boy of Set I get more marks than the top boy of Set II ? If not, how much overlap must there be ? Sometimes an artificial scale of overlapping is drawn up, as is shown in the Diagram, but the systems for overcoming this difficulty vary from school to school.

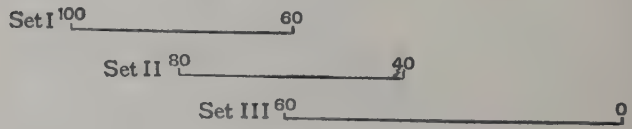

Fig. 1.

But this is by no means satisfactory.

Let us take three Forms : U. IV, M. IV and L. IV, arranged in three Latin Sets, and let us consider the case of two boys in M. IV.

A is in Set I and is at the bottom, getting 60 marks.

B is in Set II and is at the top, getting 80 marks.

So B gets more marks than A, though he is in a lower Set, and yet A has been put into Set I over B's head because he is considered to be better at Latin. A boy in Upper IV may be good in all his subjects except in Latin, and so gets placed in Set III. His

opportunities for gaining marks are restricted compared with his colleagues who are in Set I, and the boy's final position in the Form is largely prejudiced because of this one weak subject.

The question is often acute in the Second Forms in the Secondary Schools where there are some boys who have done Latin for a year and other new boys who have to be placed in a separate Set because they have not done Latin before. Those who have done Latin before get a definite advantage—not only for this reason, but because they would be placed in a higher Set carrying more marks. So the scales are weighted for some considerable time against the beginner.

There is equal objection to giving both Sets the same marks, because the boys who have done Latin before and are doing more advanced work will get only the same credit as boys who are doing the elements of the subject. The whole thing is very difficult and is the subject of much discussion amongst schoolmasters.

SCALING

In an examination paper, or during a week or fortnight's work in school, the marks allotted to a boy may not amount to a convenient number. Suppose that a boy has obtained 120 marks out of 200. Then, if the maximum mark allotted to a subject is 100, the boy's mark must obviously be divided by two, making it 60. This is a simple example of many cases which arise. In some cases the alteration of the marks brings in decimal points, and in these cases it is probably best to work to the nearest whole number, treating half a mark or more as one. Suppose a boy has obtained 73 out of a

possible 120 and the maximum allotted to that subject is 80, then the boy is credited with

$$\frac{73}{120} \text{ of } 80 = \frac{146}{3} = 48\tfrac{2}{3}.$$

Therefore 49 would be the nearest whole number. This type of scaling means multiplying by a factor. It can also readily be done by means of a graph or, better still, by a slide-rule.

In Table IIB (p. 14) it will be found that the English marks in A have been halved to produce the English marks in B. Also, since the History marks out of 140 have to be reduced to a maximum of 20, the factor for History was $\frac{1}{7}$. It will be seen from Table II that the factor for French was $\frac{1}{2}$, Latin $\frac{1}{3}$, Mathematics $\frac{1}{2}$ and Science 1. It is sometimes arranged that the top boy in a subject is scaled to get the maximum mark. Thus, in Table III, the boy F, with 51 marks in English, is scaled so that he gets the maximum of 30 marks allotted to the subject.

Similarly, in History, the boy B, with 120 marks, is scaled so that he gets the maximum 20 allotted to the subject. Similarly D with 82 in French is scaled to 50; similarly D with 91 in Latin is scaled to 50; similarly C with 98 in Mathematics is scaled to 60. Since the Science Master has been wise or lucky enough to collect his marks so that the top boy has got a maximum of 40, his marks do not require scaling.

The whole of the marks of Table II are shown in Table III, scaled so that the top boy in each subject gets the maximum allotted to the subject. The maximum marks are allotted on the basis of 10 marks per lesson per week. English has 3 lessons per week and the maximum mark is 30; History has 2 lessons per week and the maximum is 20; French has

5 lessons per week and the maximum is 50; Latin has 5 lessons per week and the maximum is 50; Mathematics has 6 lessons per week and the maximum is 60; Science has 4 lessons per week and the maximum is 40.

TABLE III

MARKS SCALED SO THAT THE TOP BOY GETS THE MAXIMUM ALLOTTED TO THE SUBJECT

Name.	English. (30)	History. (20)	French. (50)	Latin. (50)	Maths. (60)	Sc. (40)	Total.	Rank.
A	25	16	43	37	44	31	196	2
B	28	20	18	23	52	25	166	3
C	12	7	26	23	60	27	155	6
D	21	12	50	50	54	30	217	1
E	18	10	31	27	40	18	144	8
F	30	13	20	16	29	40	148	7
G	29	15	29	22	30	32	157	5
H	12	6	20	15	46	25	124	10
I	27	14	18	11	56	17	143	9
J	25	18	25	25	52	19	164	4

This scaling can be done by means of a graph, but much more quickly on a slide-rule.

In the case of the English mark, 51 on the top scale A of the rule is set over 30 on the top scale B of the slide-rule, and then all the English marks can be converted at once, for each raw mark on the rule

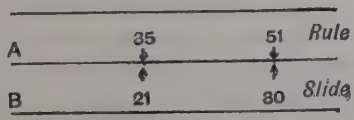

FIG. 2.

will be above its corresponding scale mark on the slide, e.g. 35 is over 21 approximately. In a similar way all the other subject marks can immediately be scaled. Thus a boy who was top in every subject would obtain 100 per cent, that is, a "possible".

Scaling also prevents any master from unduly depressing the aggregate marks. For if a master in Latin out of a maximum of 50 gives nobody more than 20, it is obvious that he is depressing the total marks obtained and also making the Latin marks of less significance in the aggregate, so that, compared with other subjects, the Latin marks are not carrying their true weight. By forcing this master to scale his top boy, who has got 20 marks, up to 50, and the other boys in the same proportion, it is ensured that the Latin should have its proper effect on the aggregate.

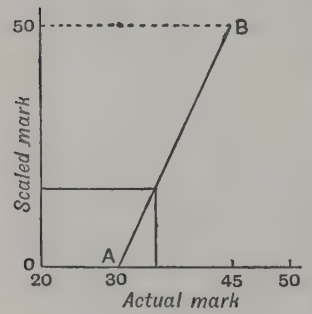

Fig. 3.

There is one other point of interest here. Suppose that in French the master's actual marks for the Form range from 30 to 45, whereas in Latin they range from 5 to 25. On scaling the top boy to 50, we get for the French a range of 33 to 50; whereas for the Latin the range is from 10 to 50. Thus the bottom boy in French is getting more marks than the bottom boy in Latin, and so to be bottom at Latin is a bigger handicap than to be bottom at French. For this reason, marks are sometimes scaled so that the top boy gets the maximum and the bottom boy 0. This is best done by means of a graph. Suppose the Latin marks range from 30 to 45. They then have to be scaled from 0 to 50. (Fig. 3.) A point A is marked corresponding to 0 on the scaled mark axis and 30 on the actual mark axis, and another point B corresponding to 50 on the scaled mark axis and 45 on the actual mark axis. The line AB is drawn,

and any actual mark can be converted into a scaled mark by reading it off on the graph. If such a system of scaling is adopted, the boy who is bottom in every subject gets a grand total of 0.

Table IV (p. 22) shows the actual cases obtained by a Form with the raw marks unscaled. At the end of the Table are two orders—the order obtained by adding up the raw marks and also the order obtained after scaling the marks to the maximum allotted to each subject.

It will be seen that there are some remarkable changes of position. This Table is given to show that the method of dealing with the marks is of the utmost importance if any reliance is to be placed on the final order.

It should be noted that two boys are equal 7th, so they are given the order $7\frac{1}{2}$ as they share the 7th and 8th places between them. Similarly, five boys compete for the 11th place, so they share the 11th, 12th, 13th, 14th, and 15th places and are given the order 13.

TABLE IV

UNSCALED

	Div.	Eng.	Hist.	Geog.	Latin.	French.	Maths.	Science.	Total.	Order.	Order after Scaling.
A .	18	100	12	16	28	83	86	42	385	12	6
B .	15	71	11	17	15	49	82	47	307	23	23
C .	19	74	14	17	48	78	62	44	356	20	16
D .	22	70	15	30	44	60	90	43	374	15	9½
E .	20	81	14	17	32	76	94	42	376	14	13
F .	14	65	15	18	25	66	90	40	333	22	20½
G .	15	89	14	21	44	78	88	44	388	11	7½
H .	16	55	11	19	45	79	74	45	344	21	18
I .	22	60	12	24	20	71	40	47	296	26	26
J .	20	81	11	23	60	70	62	38	365	18	13
K .	15	65	(12)	15	40	57	46	36	286	28	27
L .	14	53	11	19	35	51	64	46	293	27	24½
M .	13	43	10	13	15	72	46	37	249	29	29
N .	14	75	10	12	35	67	46	44	303	24	24½
O .	17	89	17	24	37	88	56	44	372	16	13
P .	17	40	15	26	69	61	34	37	299	25	28
Q .	15	81	12	17	95	68	65	38	391	10	20½
R .	14	100	8	16	84	77	60	44	403	8	22
S .	21	90	20	20	74	87	52	44	408	6	18
T .	20	89	13	25	67	78	90	45	427	3	13
U .	20	104	11	30	16	57	72	47	357	19	18
V .	20	95	21	33	37	85	84	50	425	4	2
W .	23	101	21	22	42	87	100	57	453	1	1
X .	21	90	16	24	23	81	100	44	399	9	5
Y .	22	71	11	16	24	82	94	48	368	17	9½
Z .	19	94	13	25	29	91	104	47	422	5	3
AB	16	81	21	26	26	74	111	50	405	7	4
BB	15	81	12	27	25	80	88	55	383	13	7½
CB	19	83	10	23	82	77	102	43	439	2	13
DB	16	53	4	10	15	41	40	40	219	30	30

(12) = Assessed mark.

CHAPTER II

GRAPHICAL METHODS

HISTOGRAMS AND FREQUENCY CURVES

IN dealing with marks it is often convenient to use graphical methods. The graph has a great advantage because relations can often be immediately seen which otherwise can be found only by careful and laborious examination of a Table of marks. Tabular results are often confusing, whereas in many cases a graph immediately presents a clear picture to the mind. Many instances of the use of the graph will appear in these pages and others will rapidly suggest themselves to the reader. We will now illustrate some very important graphical methods of dealing with marks.

The following Table shows the result of an Intelligence Test given to 480 boys.

TABLE V

Score Group.	Column I. Frequency.	Column II. Integrated Frequency.	Column III. Percentile.
200–180	11	480	100
179–160	65	469	98
159–140	84	404	84
139–120	106	320	67
119–100	91	214	45
99– 80	70	123	26
79– 60	42	53	11
59– 40	10	11	2
39– 20	1	1	0

The actual individual marks are not given, but the marks have been sorted out into score groups and the number of boys having marks in the respective score groups is shown in Column I. Thus from the Table it will be seen that one boy has a mark below 40, ten boys have marks between 40 and 59, 42 boys between 60 and 79 marks, and so on. There are few boys with low marks and few boys with high marks, while the greatest group is about the middle of the score scale. Thus the number of boys in the groups increases from a small number to a maximum and afterwards decreases again.

The number of boys in each group is called the "frequency", and if we draw a rectangle with the "group" as base, the height of the rectangle being the number of boys in that group, we get a figure which is called a histogram. (Fig. 4.)

If the group interval is made smaller, the histogram is made up of narrower rectangles. If the group interval is very small, then the histogram approximates to a curve with a large number of right-angle indentations.

Now in any series of measurements there will be errors due to disturbing factors which will not be under control. Hence, it would not be correct to assume that the jagged histogram is the actual state of affairs. Another series of measurements would probably give another histogram very close to the first one, and so on for each series of measurements of the same thing.

These histograms would coalesce and it would be possible to draw a curve of closest fit, which is called the frequency curve. Such a curve has been drawn in Fig. 4 by smoothing out the histogram and making the area included by the curve equal to the area of the histogram. It will be seen that the parts of the

histogram excluded by the curve are roughly compensated by the small areas included which were not in the original histogram.

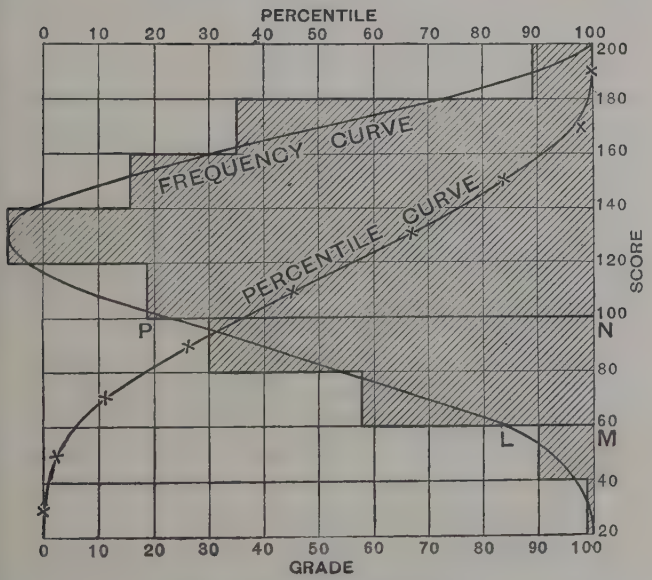

FIG. 4.—Intelligence Test 33.

COLUMN I.	COLUMN II. Integrated	COLUMN III.
Frequency.	Frequency.	Percentile.
11	480	100
65	469	98
84	404	84
106	320	67
91	214	45
70	123	26
42	53	11
10	11	2
1	1	0

It follows from the construction of the curve that the area between any two lines, such as LM and PN, must be proportional to the number of scores or measurements lying between 60 and 100, and that the

area of the whole curve will be proportional to the total number of scores, i.e. 480 in this particular case. Further, the area below PN would represent 123 scores, and the area above PN would represent (480 —123) or 357 scores.

A frequency curve can quickly be drawn by plotting the frequency against the middle of the group interval. This curve immediately suggests important conclusions, provided the number of cases under consideration is large enough. If the maximum of the curve is towards the high side, then obviously the greater number of candidates have high marks and, conversely, if the maximum is towards the low side of the mark scale, the greater number have marks below the middle score. In practice, the curve is not quite symmetrical unless a large number of cases is taken. If a test for a large number of candidates gives a symmetrical frequency curve, there is strong prima-facie evidence that the test has been a good one and has spread the candidates properly.

A frequency curve which is not symmetrical is said to be skewed. If it is skewed towards the low side, it means that marks have been difficult to get, which may be due to a high standard of marking or to the questions being too difficult. On the other hand, a skewed curve towards the upper part of the mark scale suggests that the paper has been easy or the examiners have been liberal.

If the top of the frequency curve is very flat, this indicates that the marks are very much spread out in the middle portion of the mark scale, whilst if the frequency curve has a narrow and sharp peak, it means that most of the candidates are closer together and that the paper has failed to spread the majority of the candidates out sufficiently, and that large numbers have

marks which are very close together. Some typical frequency curves are shown on page 31. In a large examination, such as the School Certificate examination, it is important for attention to be paid to the distribution of the candidates, and thus it is valuable to consider the frequency curve and its relationship to the credit line. If the credit line is anywhere near the top of the frequency curve, it means that a small movement of the credit line will affect a large number of candidates. It is better in drawing the credit line to select a position on one side or other of the central line. In practice, this line is drawn at about 45 per cent of the marks.

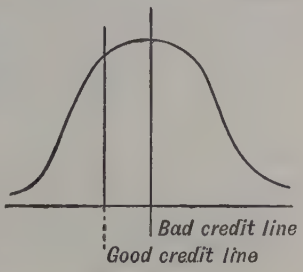

FIG. 5.

PERCENTILE CURVES

If any ordinate be drawn, as in Fig. 6, the area of the curve to the left would give the total number of scores which fall to the left of that point of the base. Thus, in Fig. 6, the area ABCD is the total number of scores which just fail to reach a mark B, and the area of the curve EFCD is the number which just fail to reach the score F.

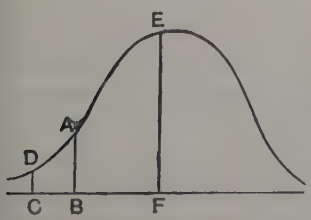

FIG. 6.

The total area up to any ordinate EF is the total number of scores or the integrated frequency up to that point. Thus, a new curve, which can be called the integrated frequency curve or percentile curve,

can be plotted. Perhaps this is best illustrated by the figures of the Intelligence Test, given in the Table on page 23. Arranging the scores in groups vertically, the first column gives the number of boys in each group interval. The second column is the integrated frequency up to each score interval.

Thus we have 1 score between 20 and 40
and 10 scores „ 40 and 60
42 „ „ 60 and 80
70 „ „ 80 and 100

and so on.

Therefore there are 11 below 60
53 „ 80
123 „ 100

The top total of the integrated frequency column is the total number of boys examined.

The totals in Column II can be reduced to a 100-point scale without impairing their relative values. This is done the most quickly by a slide-rule, and so we get :—

480 cases or 100% falling below the maximum score of 200
469 „ „ 98% „ „ a score of 180
409 „ „ 84% „ „ „ „ 160
320 „ „ 67% „ „ „ „ 140 and so on.

On plotting these percentage values against the middle of the interval, we get a percentile or a grade curve.

The curve (Fig. 4, p. 25) can be used in this way. Looking down the 50 percentile line, we see that it cuts the curve at a point corresponding to a score of 115. This means that 50 per cent of the candidates just failed to reach this score and 50 per cent of the candidates just exceeded this score. Therefore this score is called the median. Dropping a vertical line from the 25 percentile, we see that it cuts the percentile

curve at 88. This means that one-fourth of the candidates just failed to reach this score, and this is called the lower quartile or Q_2.

In the same way, dropping a vertical line from the 75 percentile, we find the score 138, which 75 per cent of the candidates just failed to reach. This score is called the upper quartile or Q_1. The scores of the middle block of candidates therefore lie between the two scores Q_1 and Q_2 or between 138 and 88. Obviously, if this distance is big, the candidates are very much spread and the average deviation of the candidates' marks from the median score is high. On the other hand, if the scores are close together, that is, if the candidates are closely grouped in the middle, then $Q_1 - Q_2$ will be small and hence the dispersion of the scores is small. $\frac{1}{2}(Q_1 - Q_2)$ is called the semi-interquartile range, and it is obviously a measure of the deviation or the spread of the scores.

If a large number of candidates are grouped together in the middle of the score scale, then naturally the semi-interquartile range is small and that group of candidates is very homogeneous.

If we take the score of any particular candidate, we can read off from the percentile curve what his percentile rank or grade is. Thus a boy with a score of 120 has a percentile rank of 55. This means that if the candidates were sorted into grades from 100 to 0, this particular candidate would be graded 55.

CHAPTER III

AVERAGES AND NORMS

THE AVERAGE

THE ordinary way to find the average score for a Form is to add up the scores and then divide by the number of boys. The figure, however, may be misleading, for one very good boy will raise the average so found and one bad boy will depress it. This is obvious from the following figures;

Case A.	Case B. (With one very good boy.)	Case C. (With one bad boy.)
83	99	83
82	82	82
80	80	80
75	75	75
68	68	68
62	62	32
6)450	6)466	6)420
Av. = 75	77·66	70

Thus the boy with the good mark raises the average by 2·66 marks, though none of the other candidates has altered his mark, and the bad boy has depressed the average by 5 marks, though the other marks are unchanged.

When we find an average, we really want to find the central tendency. Now if a large number of scores is taken, they will roughly fall on a frequency

curve, and the highest point of the curve gives the most frequent score. This is called the *mode*, which is itself a measure of the central tendency. Generally the average will not fall at the mode unless the frequency distribution is the ideal Gaussian one. (See Appendix.)

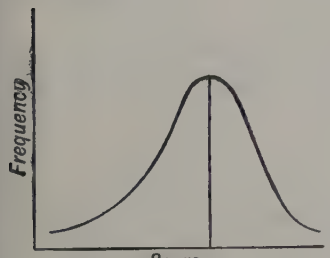

FIG. 7.

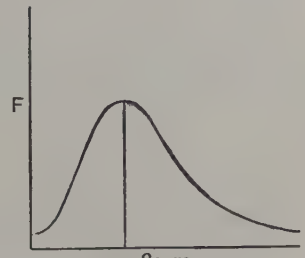

FIG. 8.

It will sometimes be found that experimental frequency curves are not symmetrical and that they are either skewed to the low side or to the high side, as shown in the figures. Fig. 7 shows one with the mode or score of most frequent occurrence on the top side of the score scale, while Fig. 8 shows the opposite result. Hence the mode obtained by inspection is not a good measure, as a rule, of the central tendency. Further, a frequency curve may be bi-modal as in Fig. 9. Such a curve is an indication that there are two separate groups amongst the candidates. The dotted lines show the distribution that might be obtained if each group were tested separately.

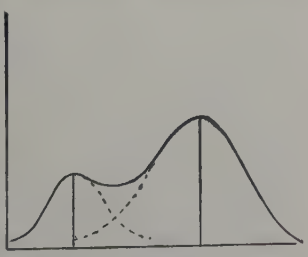

FIG. 9.

THE MEDIAN

Another method of finding the central tendency is to find the Median Score. This is the score which so divides the candidates that as many of them are above this score as are below it.

The median can often be found by inspection. In the series of scores

83, 82, 80, 79, 79, 78, 78, 78, 78, 77, 77, 76, 74, 72,

it is obvious that 5 candidates score more than 78

5 ,, ,, less ,, 78

4 ,, ,, exactly 78

The Median Score therefore is 78.

For a large number of candidates, the easiest way to find the median is to draw the percentile curve. Then the score corresponding to the 50 Percentile is the median. Further, the score corresponding to the 25 Percentile is called the Lower Quartile, and the score corresponding to the 75 Percentile is called the Upper Quartile. Thus in Fig. 4, for an Intelligence Test given to 480 candidates, the Upper Quartile Q_1 equals 138; the Median M equals 115; the Lower Quartile Q_2 equals 88.

In general, the median is the most useful measure of central tendency for ordinary school work, for individual high or low scores do not affect it as they do the ordinary arithmetical average. Caution must be observed, however, in taking the median if the number of scores is small, since one extra score may transfer the central point a considerable distance on the mark scale.

However, if an average form is given a suitable test, it will be found that the median and the average are not far apart. This is illustrated in Fig. 19, which shows the result of an Intelligence Test given to a

number of Forms. In the figure, the top mark, the bottom mark, the Upper Quartile, the median, the average (or mean) and the Lower Quartile are all shown. Interesting deductions can be made from this figure. Thus:

Form UV	has a better top half than Form MV.
Form UV	has a long tail.
Form UIII	is reasonably homogeneous.
In Form VI M.S.	the average is above the median because some good boys have got high scores.
,, Form VI Sc.	the average is below the median because some poor boys have got low scores.

NORMS

The Norm may be considered the reasonable average score, based on experience, to be expected from a group of candidates. Thus, if a large number of 12-year-old children in an Intelligence Test have scores ranging from 98 to 12, and the average score is 48, it is reasonable to say that all who have scored more than 48 are above the normal for this particular test and those below 48 are below normal for this test. We have repeated " for this test " because there is no finality about the matter unless it is found by consistent testing with a series of tests which correlate amongst themselves that a given pupil is below normal consistently. For school purposes, the easiest and most useful norm to take is the median score or the score corresponding to the 50 Percentile. This, as pointed out before, discounts the effects of a few good or bad boys.

Fig. 16 (p. 80) shows the percentile curves for a

number of boys of age groups 8 to 17 in a General Knowledge Paper. Each age group is treated separately and the percentile curves drawn as described on page 28. The scores corresponding to the 50 Percentile rank give the following norms :

Age	17	16	15	14	13	12	11	10	9	8
Norm	107	81	75	59	49	34	31	22	16	4

It should be noted that such a norm or standard score level is not a matter derived by the examiner but by the pupils themselves. It is therefore a standard of achievement—or a level. A boy's ability at this particular test can be stated not only in the marks scored, but

(1) as his Percentile rank.

Thus a score of 100 gives a Percentile rank of 80 for a boy aged 15, or

(2) as a Quotient—in this case his General Knowledge Quotient.

Thus his fellows of the same age, 15, determine the norm at 75. His score is 100. Hence his General Knowledge Quotient is $\frac{100}{75}$, or expressed as a percentage, 133 per cent.

If the test is a test of work done, it is usual to call it a Subject Quotient. Thus, if his Arithmetic marks were 50, and the norm was 41, his Arithmetic Quotient (A.Q.) is $\frac{50}{41} = 122$ per cent.

THE INTELLIGENCE QUOTIENT

A special case is that of an Intelligence Test. If his score is 120, and the norm found for boys of that age in his own school is 108, then his Intelligence Quotient (I.Q.), compared with the mean of his

fellows in that school, is $\frac{120}{108} = 1{\cdot}11$, or 111 per cent.

If the mean or norm is known for a large number of schools, then a boy can be rated against pupils of equal age in these schools. Suppose the norm over a large number of schools for this test given under same conditions is 100. Then his Intelligence Quotient compared with the greater body of pupils would be $\frac{120}{100} = 1{\cdot}2$ or 120 per cent.

The norm for a good school with carefully selected pupils is generally higher than that found by giving the same test to a large number of candidates not so selected, and so it should be noted that the Intelligence Quotient given will depend on the norm taken.

Thus for the same age group the norm for a select class of scholars in an Intelligence Test may be 130, and for the same age group the norm for a large number of secondary scholars in this Intelligence Test may be 124; also for the same age group the norm for a large number of boys and girls selected at random, irrespective of their schooling in the same Intelligence Test, may be 98. Thus the significance of the Intelligence Quotient will depend on the standard of comparison.

Schoolmasters should be cautious in making estimates of a pupil's ability and rating him as subnormal unless they measure his Intelligence Quotient in terms of the norm for the general population. For certain tests these norms can be obtained from publishers, and the Institute of Industrial Psychology publishes norms for Tests 33 and 34, which show definitely that the norm rises from Elementary to Secondary Schools.

STANDARD TESTS

Tests are now used for measuring the standard of attainment in a given subject, such as Arithmetic. Some of these have been given to large numbers of children and the norm for each age group determined. It is therefore possible to measure the standard of attainment of each pupil in terms of a level common to many schools. Each school can devise its own tests, and by giving them year by year under the same conditions, find out how the age groups in successive years compare with one another. In this way, by giving a Latin Grammar Paper two years running, it is possible to find out if one year's 15-year-age group is up to last year's similar Group.

The same method can be applied to Forms in successive years, but of course precaution must be taken to collect all the papers each year so that the test cannot be divulged. If there is any doubt, it is possible to alter the words set so as to get a test of equal difficulty without unduly endangering the efficacy of the comparison.

Some teachers fear these standard tests. Yet they should welcome them and make their own, for they give themselves thereby a footrule to measure their pupils year by year, and they are more likely to get a true judgment as to the way in which a class is taking to the work than by a general guess or comparison with a former year. The latter method is always open to various objections. Masters find that certain Forms are more difficult—often temperamentally—and this inclines them to judge erroneously. Also the work done by last year's Form, when viewed at a distance, may have that enchantment which distance certainly lends to some views.

Comparisons of Forms based on what we thought they were like last year compared with what we think we know their successors are like this year, should only be held lightly.

The comparison can easily be tested by a series of standard tests devised by the master himself, given under the same conditions and marked by himself. Then he will have some basis for judgment. Further, it may prove of the utmost value for diagnostic purposes. It often throws the work of some boys into bold relief. If the work of a boy be compared with that of his colleagues on some Standard Tests and he is below normal, yet in Intelligence Tests and General Knowledge Tests he is consistently much above normal, then that boy merits the close attention of the Formmaster and most probably of the Headmaster.

Other boys do consistently well in School, they prepare their work well, they are industrious, yet they crack up in examinations. They are called bad examinees. In many of these cases, a series of Intelligence Tests, of General Knowledge Tests, of Accuracy Tests, will often show that the boy's ability has been overrated because he is industrious, well-behaved and a neat worker. Intelligent boys who have worked consistently do not, if in a good state of health, break down in examinations as a rule. Enough has been said to show the value of careful testing for diagnostic and educational purposes.

One last word. These results should be kept and not thrown away. If they are studied year by year during a boy's school life, or for a given Form, much time and effort will be saved. For a record card is both a barometer of progress and a chart of symptoms, and the standard is an objective one. The

boys are measuring themselves by the normal standard which they have set for themselves in each age group or Form. They are not being praised or condemned by the master's standard which, after all, is subjective, liable to unknown changes and vitiated by personal equations.

Only one thing remains. The Test should be well set, the marking should be foolproof, and the conditions of the Test, such as time and place, should be followed strictly year by year.

CHAPTER IV

CORRELATIONS

CORRELATION

IF a boy is good at Arithmetic, is he likely to be good at Algebra? If he is good at Mathematics, is he less likely to be good at Latin? These questions have interested schoolmasters for a long time and many have pronounced opinions on the subject. These opinions have been based on personal observation, i.e. on the judgment of the teacher. The answer given could only be a qualitative one until recently, when the statistical device known as correlation was applied to the problem. Whereas, before, the answer was "Yes, generally speaking, a boy who is good at Arithmetic is also good at Algebra", or "No, there is no correspondence between the Arithmetic and Algebra results", we can now express the relationship in terms of a number.

If for a given set of boys the Arithmetic order is exactly the same as the Algebra order, then the correlation coefficient is said to be unity, + 1. There is obviously perfect correspondence. On the other hand, if the Algebra is exactly the reverse of the Arithmetic order, then there is inverse correspondence and the correlation coefficient is — 1. If there is no sort of relationship between the Arithmetic order and the Algebra order, and the whole thing seems to be determined by pure chance,

then there is no correlation and the coefficient is zero, 0.

Table VI (pp. 44 and 45) shows these three particular cases. In each case the rank of the boy in the subject is given, and by a comparison of these ranks we can see at once that Column I shows perfect correspondence or correlation coefficient + 1. Column II shows inverse correspondence or correlation coefficient — 1. Column III shows positive but imperfect correspondence or correlation coefficient between 0 and + 1. Column IV shows no correspondence or correlation coefficient 0. Column V shows imperfect negative correspondence or correlation coefficient between 0 and — 1.

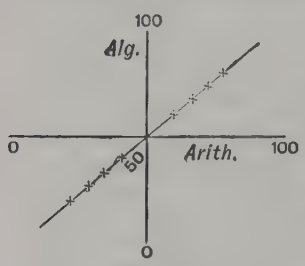

Fig. 10.

The results can also be illustrated graphically. We will take the marks to have the same range from 0 to 100 in each case and place the axes in the middle of the range. If the correspondence is perfect, the crosses which mark the positions of the boys in Arithmetic and Algebra will all be on a line making an angle of 45° with the horizontal axis. The correlation coefficient is + 1 or tan 45°, or for any boy the ratio $\frac{\text{Algebra Mark}}{\text{Arithmetic Mark}} = 1$ (Fig. 10). If the correspondence is inverse we get Fig. 11, and there the line of relationship has a back-hand slope and the coefficient of correlation is — 1. If there is no sort of correspondence, the crosses will be dotted about here and there by pure chance, and obviously it is impossible to predict from a boy's Arithmetic position

what his position in Algebra will be. There is therefore no correlation (Fig. 12). Now there may be a partial relationship between the Arithmetic order and the Algebra order, which would be expressed by a

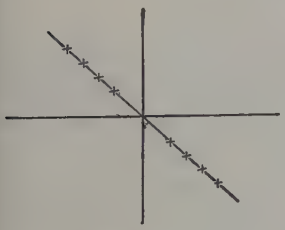

Fig. 11.

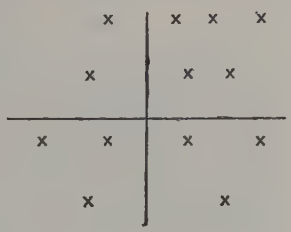

Fig. 12.

correlation coefficient between 0 and 1. To show this for any given scores in Arithmetic and Algebra, it is first of all necessary to adjust the scales of the axis so as to make the standard deviations equal. For our purpose the same thing will be arrived at by choosing scales so that the interquartile range which comprises the middle half of the marks is represented by the same length on each axis. If this is done, and it is found that the crosses be nearly on a line which can be represented by PP′, then the correlation coefficient is the slope of this line PP′ (Fig. 13). The

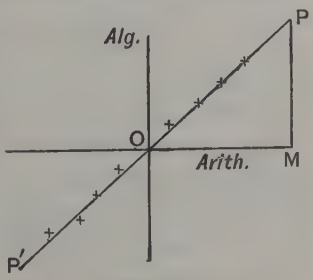

Fig. 13.

slope of such a line is measured by tan POM, or $\frac{PM}{OM}$.

If the ratio is 0·6, then the correlation coefficient r is + 0·6. Such a line is called a line of regression, and the slope of the line of regression is the correlation

coefficient. In practice, the crosses will rarely fall on a line, but the line of nearest fit can be drawn, and the slope of this is then a rough measure of the relationship between the two tests.

THE METHOD OF RANKS

Fortunately, it is not necessary to draw the graphs or to bother about adjusting the interquartile range, for the coefficient can be calculated direct by what is known as the method of ranks. This method has been developed by Spearman and consists in using the rank or order of a boy in his Form for a given test. The absolute marks are disregarded except in so far as they determine each boy's rank. In Table VI, Column III, let us suppose we have the Arithmetic and Algebra orders for a number of boys. The difference of rank, d, is found by subtraction and placed in the third column, and the fourth column gives the square of the rank differences, d^2.

All these squared rank differences will be positive, whether a boy loses or gains in rank in passing from one test to the other, and the sum of these squares, Σd^2, is a measure of the disagreement to be found between the Algebra and Arithmetic orders.

Now if there are N boys, it can be shown by the Mathematical Theory of Probability that the sum of the differences of the squared ranks which could arise by pure chance is $\frac{1}{6}N(N^2 - 1)$.

The ratio K =

$$\frac{\text{Sum of actual rank differences squared}}{\text{Sum of rank differences squared to be expected by chance}}$$

is a measure of the disagreement of the results. Hence $(1 - K)$ is a measure of the agreement of the results and $(1 - K)$ is akin to the correlation coefficient since

it is a measure of agreement expressed in quantitive terms, and we shall denote it by ϱ.

The Table VI (p. 44, 45 and 46) showing.

Perfect Correlation
No Correlation
Negative Correlation
Positive Imperfect Correlation

should now be studied.

In the cases set out N is small, but the method is exactly the same for bigger numbers. This method can be used profitably up to 30, the usual size of a form in a secondary school. It may be found that one or two boys exercise a great effect on the coefficient by their change of rank. Sometimes, in practice, one finds fair correlation for twenty-nine boys in a Form, but the whole coefficient is very much altered by the thirtieth case.

The value of the factor $\frac{1}{6}N(N^2 - 1)$, or

$$\tfrac{1}{6}(N - 1)N(N + 1),$$

can be calculated and tabulated for the numbers of boys usually to be found in a Form, and so the work reduces itself in any given case mainly to finding the rank differences squared, i.e. Σd^2.

The coefficient ϱ, given by the method of ranks, is not quite the same thing as calculated from the line of regression, but the relationship is close.

It is shown in more advanced books that $r = 2\sin\frac{\pi}{6}\varrho$

and a Table of Corrections can be obtained. The correspondence is so close that for all practical purposes the correction can be neglected. It is probably within the error of experiment of marking, for examiners and teachers are only human and liable to err.

TABLE VI

Column I.

Name.	Test 1 Rank.	Test 2 Rank.	Rank Difference d.	d^2.
A	1	1	0	0
B	2	2	0	0
C	3	3	0	0
D	4	4	0	0
E	5	5	0	0
F	6	6	0	0
G	7	7	0	0
H	8	8	0	0
I	9	9	0	0
J	10	10	0	0

$\Sigma d^2 = 0.$

Hence this is perfect correlation.

Column II.

Name.	Test 1 Rank.	Test 2 Rank.	Rank Difference d.	d^2.
A	1	10	9	81
B	2	9	7	49
C	3	8	5	25
D	4	7	3	9
E	5	6	1	1
F	6	5	1	1
G	7	4	3	9
H	8	3	5	25
I	9	2	7	49
J	10	1	9	81
				330

$\therefore \Sigma d^2 = 330.$

As $N = 10$, $\frac{1}{6}N(N^2 - 1) = 165.$

$\therefore \frac{330}{165} = 2.$

$\therefore$ Corr. Coeff. $= 1 - 2$

$= -1$

i.e. Inverse or Negative Correlation.

Column III.

Name.	Test 1 Rank.	Test 2 Rank.	Rank Difference d.	d^2.
A	1	4	3	9
B	2	7	5	25
C	3	1	2	4
D	4	8	4	16
E	5	5	0	0
F	6	6	0	0
G	7	2	5	25
H	8	3	5	25
I	9	9	0	0
J	10	10	0	0
				104

$$\therefore \Sigma d^2 = 104$$

$$\text{Corr. coeff.} = 1 - \frac{\Sigma d^2}{\frac{1}{6}N(N^2-1)}$$

$$= 1 - \tfrac{104}{165}$$

$$= 1 - \cdot 63$$

$$= \cdot 37$$

i.e. positive but imperfect correlation.

Column IV.

Name.	Test 1 Rank.	Test 2 Rank.	Rank Difference d.	d^2.
A	1	6	5	25
B	2	3	1	1
C	3	2	1	1
D	4	7	3	9
E	5	10	5	25
F	6	5	1	1
G	7	9	2	4
H	8	4	4	16
I	9	8	1	1
J	10	1	9	81
				164

$$\therefore \Sigma d^2 = 164$$

$$\text{Corr. coeff.} = 1 - \frac{\Sigma d^2}{165}$$

$$= 1 - \tfrac{164}{165}$$

$$= 1 - 1$$

$$= 0$$

$\therefore$ there is no correlation at all.

COLUMN V.

Name.	Test 1 Rank.	Test 2 Rank.	Rank Difference d.	d^2.
A	1	9	8	64
B	2	7	5	25
C	3	5	2	4
D	4	2	2	4
E	5	4	1	1
F	6	8	2	4
G	7	3	4	16
H	8	6	2	4
I	9	10	1	1
J	10	1	9	81
				204

$$\Sigma d^2 = 204$$

$$\therefore \text{ corr. coeff.} = 1 - \frac{\Sigma d^2}{165}$$

$$= 1 - \tfrac{204}{165}$$

$$= 1 - 1{\cdot}236$$

$$= \quad - \ {\cdot}236$$

or imperfect negative correlation.

Thus the correction appears to be an unnecessary refinement which practical schoolmasters need not bother about. The following Table gives point to this. It seems hardly necessary to stake our reputation on the figures in the second and third decimal places.

ϱ . . .	0·50	0·55	0·60	0·65	0·70	0·75	0·80
$r = 2 \sin \frac{\pi}{6} \varrho$	0·5176	0·5680	0·6180	0·6676	0·7167	0·7654	0·8135

A full Table is given in Rugg's *Statistical Methods applied to Education.*

THE FOOTRULE

Spearman has devised another formula which he calls the "Footrule" formula. In this attention is paid only to the *gains* of rank. These are added up and losses of rank are neglected. Then if Σg = sum of gains of rank,

the coefficient of agreement $R = 1 - \frac{\Sigma g}{\frac{1}{6}(N^2 - 1)}$

This formula is easy to use, but is not so accurate as the "squared difference" method. It is, however, not intended to give precise results, but only to show whether there is a measure of agreement.

The coefficient R can be corrected into r, the correlation coefficient, by means of the formula

$$r = 2 \cos \frac{\pi}{3}(1 - R) - 1,$$

but this is generally unnecessary, for the Footrule Formula is only a rough method of showing relationship and is not intended to be an exact statistical measure. The following Table shows the magnitude of the correction.

Table.

R	·1	·3	·5	·7	·9
r	·17	·48	·73	·90	·99

Example: Footrule method.

Let us take Column III, p. 45, and work it out by this method.

		Gains in rank.	
A	1	4	
B	2	7	
C	3	1	+ 2
D	4	8	
E	5	5	
F	6	6	
G	7	2	+ 5
H	8	3	+ 5
I	9	9	
J	10	10	
			12

$\therefore \Sigma g$ = sum of gains in rank

$= 12$

$$\text{Corr. coeff.} = 1 - \frac{\Sigma g}{\frac{1}{6}(N^2 - 1)}$$

$$= 1 - \frac{12}{\frac{1}{6}(100 - 1)}$$

$= 1 - \frac{72}{99} = 1 - \cdot 73$

$R = 0 \cdot 27$ or $r = \cdot 44$

The Probable Error is $\frac{\cdot 43}{\sqrt{10}} = \cdot 14$

The method of rank differences squared gave 0·37.

It can be shown that the Probable Error (P.E.) of a rank correlation coefficient (ϱ) is

$$0 \cdot 706 \frac{(1 - \varrho^2)}{\sqrt{N}}$$ (see page 54).

In the above case the P.E. = ·19. Since in each case the P.E. is about half the correlation coefficient, it shows that neither coefficient is very reliable. The greater the number (N) of cases, the smaller the P.E. and hence the greater is the reliability of the correlation coefficient.

THE PRODUCT-MOMENT METHOD.

Let a boy score X in a test when the average score of the Form is a.

Let his distance from the average be x which equals $(X - a)$.

Let the same boy score Y in a second test where the average is b.

Let his distance from the average be Y so that $Y = (Y - b)$.

Tabulate X, Y, x^2, y^2 and xy for each boy.

Then summing up and paying attention to sign,

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \,.\, \Sigma y^2}}$$

Example.

From the Table, it is easily seen that $a = 11$ & $b = 5$. Hence $x = X - 11$ & $y = Y - 5$.

Name.	X	Y	x	y	x^2	y^2	xy
A . .	20	7	9	2	81	4	18
B . .	18	5	7	0	49	0	0
C . .	16	9	5	4	25	16	20
D . .	14	6	3	1	9	1	3
E . .	12	7	1	2	1	4	2
F . .	10	4	−1	−1	1	1	1
G . .	8	6	−3	+1	9	1	−3
H . .	6	3	−5	−2	25	4	10
I . .	4	2	−7	−3	49	9	21
J . .	2	1	−9	−4	81	16	36
Totals .	110	50	—	—	330	56	108
Average	11	5	—	—	—	—	—

$$\therefore\ r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \,.\, \Sigma y^2}} = \frac{108}{\sqrt{330 \times 56}} = \cdot 79.$$

The Table merely illustrates the arrangement of the work and it should be understood that the results are not very reliable with so few as 10 cases.

The product-moment method is most suitable for cases exceeding 30. Where the number of cases is 30 or less, it is generally sufficient to use the method of ranks.

PROBABLE ERROR (P.E.) OR PROBABLE DEVIATION

In a curve of normal frequency, it has been shown that the frequencies will group themselves about a mean position which is the mean score M.

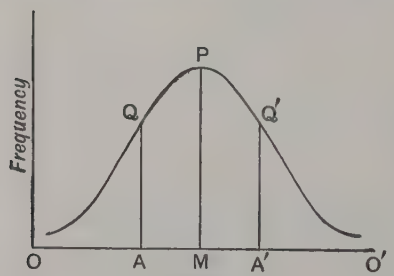

FIG. 14.

The area on the left-hand side of PM will include *half* of the scores and the area on the right-hand side of PM will include the other *half* of the scores. Now the total number of cases is represented by the *area* between the curve and the base. If a point A is found such that the line AQ divides the left-hand area into two equal parts, and a point A′ is found such that the line A′Q′ divides the right-hand half into two equal parts, it is obvious that the lines AQ, PM, A′Q′ divide the total area into four equal parts:

Then	Area	OAQ	represents	$\frac{1}{4}$	of	the	scores.
,,	,,	QAMP	,,	$\frac{1}{4}$	,,	,,	,,
,,	,,	PMA′Q′	,,	$\frac{1}{4}$	,,	,,	,,
,,	,,	Q′A′O′	,,	$\frac{1}{4}$	,,	,,	,,
And	,,	QAA′Q′	,,	$\frac{1}{2}$	,,	,,	,,

Therefore $\frac{1}{4}$ of the scores lie between score A and score M
and $\frac{1}{4}$ „ „ „ „ „ „ M „ „ A′
and $\frac{1}{2}$ „ „ „ „ „ „ A „ „ A′

Consider the areas OAQ and QAMP. Since a *quarter* of the scores lie between A and M, all this particular group deviate from the mean score by an amount which is less than AM. Since a *quarter* of the scores lie between O and A, all this particular group deviate from the mean score M by an amount greater than AM. Hence because $\frac{1}{4}$ of the scores deviate by less than AM and $\frac{1}{4}$ „ „ „ „ „ more than AM, the *probable deviation* for the group of *half* the scores represented by the left-hand side of the curve is AM.

In the same way, it can be shown that for the right-hand part of the curve, the top half of the scores is greater than the mean score M by an amount which is approximately MA′. Hence, for a large number of scores with various deviations from the mean score M, the *probable deviation* is either AM below or MA′ above, i.e. by an amount $\pm$ AM, for AM = MA′ in a normal curve.

Now it has become the practice to call this the *probable error* and not the probable deviation. This grew up because the standard curve was once known as the curve of errors. If AM = MA′ = P.E. (the probable error), we say that the mean score is M, with a probable error of $\pm$ P.E. The P.E. is therefore the most probable deviation from the mean, and it is also the deviation which is just as likely to be exceeded as not to be exceeded.

For a normal curve the P.E. is obviously the quartile deviation or the semi-interquartile range, i.e. $\frac{Q_1 - Q_2}{2}$ and is called Q. For such a normal curve, Q_1 will be just as much above the median as Q_2 is

below it. It is probably better to keep the term P.E. for distributions which follow the normal curve.

As most educational data are obtained from small samples of the school population, the frequency curve is rarely quite symmetrical and is generally skewed. Hence for such curves, the probable deviation can best be expressed in terms of the semi-interquartile range, i.e. $\frac{Q_1 - Q_2}{2}$ = the quartile deviation Q. In such a skewed distribution, the median will not be halfway between the quartile points.

To sum up.

The Upper Quartile Q_1 is the score corresponding to the 75 Percentile.

The Lower Quartile Q_2 is the score corresponding to the 25 Percentile.

The median is the score at the 50 Percentile. Hence, the probable deviation from the median score will be

$$\pm \frac{(Q_1 - Q_2)}{2}$$

e.g. suppose score for 75 Percentile is 64
,, ,, ,, 25 ,, ,, 30
,, ,, ,, 50 ,, ,, 48

then the median score is 48 and the probable deviation

$$\text{is } \pm \tfrac{1}{2}(64 - 30)$$
$$= \pm 17.$$

Therefore half of the total scores will be within a range of $\pm$ 17 from the median score 48.

COEFFICIENT OF VARIABILITY

We have seen that the median (M) can be taken to be a fair measure of the average score, that the probable deviation (or the P.E.) is a measure of the

dispersion of the scores about this central tendency, and for an experimental curve, the semi-interquartile range Q can be taken as the probable deviation.

Now the relation which the probable deviation bears to the median score is of importance, for it measures the degree of the variability of the results. Hence, the ratio $\frac{Q}{M}$ is a measure of the variability.

This can be expressed as a percentage by multiplying by 100, and then $\frac{100Q}{M}$ is called the coefficient of variability, V. This is independent of the actual scores or marks obtained, and expresses the relative dispersion.

If V is greater than 25 per cent, it means that the spread is rather big and the results should be treated with caution.

PROBABLE ERROR OF CORRELATION COEFFICIENT

When we were finding the correlation coefficient between Arithmetic and Algebra, we were dealing with a small number of children. Suppose this coefficient is 0·6. Would it still be 0·6 if we took another group of children? It probably would not. If a large number of groups were taken, it would be found that the correlation coefficient would vary around a certain mean. Some would give smaller results and some larger results, and so we could plot the frequency curve. (Fig. 15.)

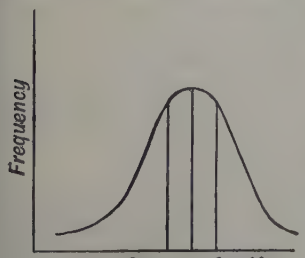

FIG. 15.

Let us suppose the mean value of all these correlation coefficients is 0·62 and that one-half of the results lie at a distance of 0·03 on either side of this mean. We could then say that the correlation coefficient is 0·62, with a probable deviation of 0·03. If the curve is approximately normal, the probable error or P.E. $= \pm$ ·03.

In measuring a correlation coefficient, it is obvious that the fewer the number of cases, the larger the probable deviation or probable error. The P.E. is really a measure of the unreliability of the results. If the P.E. is big compared with the coefficient, the results are obviously unreliable. If r is the correlation coefficient, then results should always be stated in the form $r \pm$ P.E. If r is not greater than 3 P.E., the results are generally so unreliable as to be treated with extreme caution. In fact, some observers discard any results when the P.E. is greater than *one quarter* of the correlation coefficient.

Calculation shows that the P.E. in a Coefficient r obtained from a group of number N is given by

$$\text{P.E.} = \frac{\cdot 706(1 - r^2)}{\sqrt{\text{N}}}$$

Thus, if the value of r for 15 cases is 0·86,

$$\text{the probable error P.E.} = \frac{\pm \cdot 706\,(1 - \cdot 86^2)}{\sqrt{15}} = \pm \cdot 047$$

$$\therefore r = \cdot 86 \pm \cdot 047.$$

It should be noted that r has been used here, and if the coefficient ϱ is found by the Spearman method of the rank differences squared, we have seen that

$$r = 2 \sin \left(\frac{\pi}{6}\varrho\right).$$

Between 0 and + 1, this formula always gives $r > \varrho$ by a small amount.

Now $(1 - \varrho^2) > (1 - r^2)$. Hence, substituting $(1 - \varrho^2)$ for $(1 - r^2)$ increases the value of P.E. obtained. Therefore, if we use a value of ϱ in the formula instead of r, we are not unwittingly pretending our results are more reliable. We are, in fact, increasing the unreliability slightly. Hence if ϱ is taken, the P.E. stated will be on the top side, which is better than under-estimating.

[This can be shown by the calculus.

$$\text{Let P.E.} = e;$$

$$e = \frac{0{\cdot}706}{\sqrt{N}}(1 - r^2) = A(1 - r^2)$$

$$de = -2Ardr.$$

Hence e will be too big if r is too small. If we take ϱ for r, we are assessing r too small and hence we are making the Probable Error too big.]

SUMMARY OF P.E. FOR CORRELATION COEFFICIENTS

If r is the correlation coefficient from the line of regression or product-moment method, the P.E. $= 0{\cdot}6745 \frac{1 - r^2}{\sqrt{N}}$.

If ϱ is the correlation coefficient found by "rank difference squared" method, the P.E. $= 0{\cdot}706\frac{1 - \varrho^2}{\sqrt{N}}$.

If R is the correlation coefficient found by "the gains of rank" method, the P.E. $= \frac{0{\cdot}43}{\sqrt{N}}$.

TABLE VII

Name.	Intelligence.	Order.	General Knowledge.	Order.	d.	d^2.
A	106½	13	111	7	6	36
B	110½	10	83	20½	10½	110¼
C	121½	6	105	12	6	36
D	111	9	106	11	2	4
E	133½	3	125	2½	½	¼
F	86	23	107	10	13	169
G	82	25	71	25	0	0
H	97½	18	83	20½	2½	6¼
I	91	20	95	17	3	9
J	86½	22	99	15	7	49
K	126	4	101	14	10	100
L	94½	19	57	26	7	49
M	101	16	77	22	6	36
N	134½	2	102	13	11	121
O	116½	7	140	1	6	36
P	109½	11	119	4	7	49
Q	106	14½	87	19	4½	20¼
R	106	14½	110	8½	6	36
S	122	5	112	6	1	1
T	160	1	125	2½	1½	2¼
U	107	12	115	5	7	49
V	113	8	110	8½	½	¼
W	—	—	—	—	—	—
X	88	21	94	18	3	9
Y	—	—	—	—	—	—
Z	78½	26	76	23	3	9
AB	98	17	97	16	1	1
BB	82½	24	73	24	0	0
						938½

Corr. Coeff. = 0·68.
P.E. = ± 0·075.

SOME EXAMPLES OF CORRELATION

Table VII shows a Form List, giving the marks in an Intelligence Test and a General Knowledge Paper. From the marks the ranks in each test were found, and are shown in the column next to the marks. It is important that these ranks should be assessed properly. The difficulty arises when two or more boys are equal.

If two boys were equal, 14th, it is usual to mark them = 14 and then the next boy becomes the 16th. For correlation purposes this will not do. The boys are then marked as of rank $14\frac{1}{2}$, for they share the 14th and 15th places between them. An example of this occurs in Table VII.

Similarly, if three boys are placed together in the 10th place, they share the 10th, 11th and 12th places, i.e. their ranks will be 11th.

Turning to Table VII (p. 56), we see set out the square of the rank differences which are totalled to give $938\frac{1}{2}$. Since there are twenty-six boys in the Form, this must be divided by a factor $F = \frac{1}{6}N(N^2 - 1)$. The following Table may be useful in saving time when a number of correlations have to be worked.

N =	25	26	27	28	29	30	31
F =	2,600	2,925	3,276	3,654	4,060	4,495	4,960

In this case $N = 26$. Hence $F = 2{,}925$.

$$\text{Therefore } \frac{\Sigma d^2}{F} = \frac{938\frac{1}{2}}{2{,}925} = 0{\cdot}32.$$

Then the correlation coefficient $= 1 - 0{\cdot}32$
$= \underline{\underline{0{\cdot}68}}$

$$\text{The P.E.} = \pm \frac{0{\cdot}706 \times (1 - 0{\cdot}68^2)}{\sqrt{26}} = \pm \underline{\underline{0{\cdot}075.}}$$

RELIABILITY OF A TEST

The results of a test with the same Form will not always be the same, owing to the variations in the boys themselves which may be produced by health or temperament. We all have our "off" days. Yet, despite these difficulties, a good test will give roughly the same results on all occasions. The top boys are generally top and the bottom boys at the bottom. If a test satisfies this it is said to be reliable. If a test does not produce consistent results, then it has some inherent weakness of construction or marking and so is unreliable. Obviously, the reliability of a test can only be found by giving the test to the same pupils several times and comparing the results. If the results are consistent, then the test is reliable.

The reliability or consistency may be gauged by correlating the results. Though the scores may increase, owing to practice and experience, this factor is approximately the same for all the members of the Form, and so the orders or ranks will not be materially distorted. Then the correlation coefficient will be high. Therefore, to ascertain the reliability of a test, it should be given several times and the results correlated by the method of ranks.

Sometimes, to prevent the same test being used several times, duplicate tests which are very similar, but with a change of words, are used. If the reli-

EXAMPLE OF TEST FOR RELIABILITY

TABLE VIII

Name.	Score.	Rank.	Percent. Increase.	Score.	Rank.
A	143	3	20	171½	1
B	119½	15	14	137	18
C	123½	11½	22	151	8
D	125½	10	17	147½	14
E	137	4	21	166	2
F	—	—	—	—	—
G	106½	20	24	132½	24
H	99	23	32	130½	26
I	121½	14	22	148	12½
J	117	16½	30	152	7
K	111	18	24	137½	16½
L	133½	6	17	156	4
M	102	22	33	136	19½
N	145½	2	6	154	6
O	109	19	22	133	22½
P	86	26	56	134	21
Q	97½	24	36	133	22½
R	106	21	24	132	25
S	91	25	49	136	19½
T	123½	11½	20	148	12½
U	126	9	19	150½	9
V	134½	5	16	155½	5
W	123	13	12	138	15
X	117	16½	17	137½	16½
Y	160	1	3	165½	3
Z	132½	7	12	149	10
AB	78½	27	40	110	27
BB	126½	8	17	148	12½
Total	3,195½	—	625	3,889½	—
Average	118·3	—	23	144	—

$r = 0{\cdot}90.$

P.E. $= \pm 0{\cdot}026.$

ability of the test is low, then the test needs reconstruction or lengthening. No test can be considered satisfactory until it gives consistent results as shown by the fact that the coefficient of reliability is high. Some writers suggest it should be as high as 0·9.

A good test should also, as far as possible, eliminate the personal equation of the examiner. If the same test is given to the same Form by two different examiners and the correlation is high, then the test is an objective one. If the correlation is low, the results obviously depend on the examiners and the test is a subjective one. All true tests should have a high objectivity so that the results will be independent of the personality of the marker.

In this case the same Intelligence Test was given to the same boys two years running. The coefficient of correlation between the results for the Form shown on Table VIII was good, being $0{\cdot}9 \pm 0{\cdot}026$. This means that though most boys have increased their marks, owing to practice and mental growth, the order of intelligence is about the same. So whatever the Intelligence Test measured, it apparently measured the same thing each year. Hence the Test was reliable for this Form.

The coefficient was worked out for all the Forms taking the same test two years running, and the results obtained were as follows: 0·85, 0·82, 0·80, 0·55, 0·54, 0·70, 0·78, 0·90, 0·42, 0·70, 0·60, 0·71, 0·48, 0·33, 0·67. The average value is 0·66.

Considering the large number of boys and Forms tested, this is excellent evidence of the reliability of the Test. The test used was Test No. 33 of the National Institute of Industrial Psychology.

ROUGH METHODS OF FINDING RELATIONSHIP

Pearson's Method.

This consists in making a fourfold Table.

If a number of boys are tested in Essay and Arithmetic, then they may be divided into 4 classes :

In the fourfold Table, above average = H, below average = L.

a = Number above the average in Essay and Arithmetic, i.e. H.H.

b = Number above the average in Essay and below average in Arithmetic, i.e. H.L.

c = Number below the average in Essay and above average in Arithmetic, i.e. L.H.

d = Number below the average in Essay and Arithmetic, i.e. L.L.

Then Pearson's Coefficient is

$$p = \cos\left(\frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}\right)\pi.$$

	Essay	
Arithmetic	H.H. a	H.L. b
	L.H. c	L.L. d

In an actual case for 36 Scholarship Candidates in Essay and Arithmetic, the figures were

H.H. = 10 H.L. = 8
L.H. = 8 L.L. = 10

10	8
8	10

By Pearson's Formula,

$$p = \cos\left(\frac{\sqrt{64}}{\sqrt{100}+\sqrt{64}}\right)\pi$$

$$= \cos\frac{8}{18}\pi = \cos\frac{6\pi}{9} = 0{\cdot}17$$

Sheppard's Method of Unlike Signs.

Sheppard has suggested a modification of Pearson's formula for roughly measuring relationship.

The coefficient s is given by $s = \cos\frac{U}{L+U}\pi$, where

U = percentage of cases of unlike signs (HL + LH),
L = percentage of cases of like signs (HH + LL).

Obviously U + L = 100.

Hence, $s = \cos\frac{U \times 180}{100}$, putting $\pi = 180°$

$$= \cos 1{\cdot}8U°.$$

It is clear that U must lie between 50 and 0 for positive correlation and between 50 and 100 for negative correlation. A Table can be constructed for each value of U so that s can be read right off.

In the above cases, since H.L. = 8 and L.H. = 8, (H.L. + L.H.) = 16 out of 36 cases.

a	b
c	d

$$\therefore U = \frac{16}{36} \times 100 = 44 \text{ per cent}$$

$$\therefore s = \cos (1{\cdot}8 \times 44)^\circ = \underline{\underline{0{\cdot}1873}}$$

This is in fairly good agreement with Pearson's coefficient.

COEFFICIENT OF ASSOCIATION

For a fourfold Table, a further coefficient, measuring the intensity of the association between two series of data, is called the coefficient of association Q, and is given by

$$Q = \frac{ad - bc}{ad + bc}.$$

For the above case, $Q = \dfrac{10 \times 10 - 8 \times 8}{10 \times 10 + 8 \times 8}$

$$= \frac{100 - 64}{100 + 64} = \frac{36}{164} = 0{\cdot}22.$$

COEFFICIENT OF COLLIGATION

A modification of this coefficient is called the coefficient of colligation and is given by

$$\omega = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}.$$

This is sometimes more cumbersome to deal with,

but it gives more conservative estimates of relationship than the coefficient of association Q.

In the above case of Essay and Arithmetic,

$$\omega = \frac{\sqrt{100} - \sqrt{64}}{\sqrt{100} + \sqrt{64}} = \frac{10 - 8}{10 + 8} = \frac{2}{18} = 0{\cdot}1111$$

To sum up, the relationship between the Essay and Arithmetic results is as follows:

Pearson's Coefficient . .	$p = 0{\cdot}17$
Sheppard's	$s = 0{\cdot}19$
Coefficient of Association . .	$Q = 0{\cdot}22$
Coefficient of Colligation . .	$\omega = 0{\cdot}11$

It should be noted that ω gives the lowest value.

Correlation by the method of ranks for the same case gives $\varrho = 0{\cdot}36$, with a Probable Error of

$$\frac{0{\cdot}706(1 - 0{\cdot}36^2)}{\sqrt{36}} = 0{\cdot}10.$$

$$\therefore r = 0{\cdot}36 \pm 0{\cdot}10.$$

For the same 36 boys the relationship between Essay and Intelligence Test was as follows:

H.H. = 10 L.H. = 8
H.L. = 8 L.L. = 10

10	8
8	10

These give the same coefficients as for the Essay and Arithmetic.

The method of ranks gave $0{\cdot}46 \pm 0{\cdot}093$.

For English and History the same results were found.

The method of ranks gave $0{\cdot}44 \pm 0{\cdot}095$.

For History and Scripture the figures were

H.H. = 12 L.H. = 6
H.L. = 6 L.L. = 12

12	6
6	12

Hence

$p = 0{\cdot}50$
$s = 0{\cdot}51$
$Q = 0{\cdot}60$
$\omega = 0{\cdot}33$

Method of ranks gave $0{\cdot}44 \pm 0{\cdot}095$.

For Essay and English.

H.H. = 13 L.H. = 5
H.L. = 5 L.L. = 13

13	5
5	13

Hence

$p = {\cdot}64$
$s = {\cdot}64$
$Q = {\cdot}74$
$\omega = {\cdot}44$

Method of ranks gave $0{\cdot}50 \pm 0{\cdot}088$.

Exactly similar results were obtained for the Arithmetic and Intelligence Test, showing that there was some positive correlation in this case. Method of ranks gave $0{\cdot}31 \pm 0{\cdot}10$.

SUMMARY OF RESULTS

Subjects.	p.	s.	Q.	ω.	ρ.
Essay and Arithmetic	0·17	0·19	0·22	0·11	$0{\cdot}36 \pm 0{\cdot}10$
Essay and Intelligence	0·17	0·19	0·22	0·11	$0{\cdot}46 \pm 0{\cdot}093$
English and History .	0·17	0·19	0·22	0·11	$0{\cdot}44 \pm 0{\cdot}095$
History and Scripture	0·50	0·51	0·60	0·33	$0{\cdot}44 \pm 0{\cdot}095$
Essay and English .	0·64	0·64	0·74	0·44	$0{\cdot}50 \pm 0{\cdot}088$
Arithmetic and Intelligence	0·64	0·64	0·74	0·44	$0{\cdot}31 \pm 0{\cdot}10$

With regard to the last correlation of Intelligence Test and Arithmetic, the coefficient was much affected by the startling change of place in the case of three boys. If these three boys were neglected, then the correlation coefficient for the other thirty-three turned out to be 0·42.

With regard to the above values, the most reliable is probably ρ. The others are given for comparison purposes only, and it should not be assumed that all the coefficients are of equal validity as a measure of correlation.

It will be seen that Q is generally fairly high, while ω gives a more conservative estimate of the correspondence. For this reason the coefficient of colligation ω is much to be preferred to the coefficient of association Q, which, by its tendency to a high value, may be misleading.

CHAPTER V

MANAGEMENT OF SCHOOL MARKS

RETURNS

WHEN a large number of marks have to be handled, it is convenient to have the returns from the Forms arranged on some fixed scheme or method. If this is not done and the marks for 600 boys in a General Knowledge paper are sent in on all sorts of forms, the work of arranging them in age groups and score groups becomes very laborious and errors are difficult to check. A convenient form of return is given on page 68, Table IX.

On this Form List it will be seen that the marks are arranged in parallel columns according to the age groups and that the age of each boy is given at the side. In this book, whenever an age group is referred to, it is the age determined on the previous July 31st which is meant. It is also convenient for purposes of correlation to have the ranks given as shown on the list.

FORM E.

TABLE IX

GENERAL PAPER

Age.	Name.	Rank.	Age 13.	Age 14.	Age 15.	
15.0	A .	8			109	
14.11	B .	3		124		
15.9	C .	2			128	
15.5	D .	13			95	
14.8	E .	1		135		
15.9	F .	10½			105	
14.10	G .	18½		80		
14.11	H .	16		90		
15.10	I . .	4			118	
14.1	J .	21		74		
13.5	K .	6	114			
14.3	L .	24		67		
14.3	M .	18½		80		
13.9	N .	12	104			
14.9	O .	9		108		
13.2	P .	23	69			
14.8	Q .	16		90		
15.2	R .	10½			105	
15.9	S .	20			79	
14.7	T .	14		92		
15.0	U .	7			111	
14.9	V .	22		72		
14.4	W .	16		90		
15.0	X .	5			117	
		Totals	287	1,102	967	=2,356=Av. 98·1
		No. N.	3	12	9	= 24

. . . . . .

The following General Paper was set to all Forms and the returns in the form shown above were collected ready for analysis as shown in Tables X and XI.

Name.....

Form.....

Age on July 31*st*, 1927.....

GENERAL PAPER, 1928.

Time—1 *Hour.*

All answers to be written opposite the questions on the RIGHT.

Write at the top your Name, Form and Age.

Total Marks—250.

1.—For what is each of the following famous ?
Ibsen.
Wren.
Byrd.
Rude.
Velasquez.

2.—To what countries do the following Islands belong ?
Spitzbergen.
The Seychelles Group.
Zeeland.
Rhode Island.
Rhodes.

3.—
Which is the largest planet ?
Which planet is nearest to the sun ?
Which planet has rings ?
Which planet takes the longest to go round the sun ?
Who discovered the planet Uranus ?

4.—Complete the following quotations :—
A little —— is a dangerous thing.
To-morrow to fresh —— and pastures new.
Not once or twice in our fair island story
The path of —— was the way to glory.
I could not love thee, dear, so much,
Loved I not —— more.
They also —— who only stand and wait

5.—Who was :—
The Lady of the Lamp.
The Man of Destiny.
The Lion of the North.
Old Noll.
The Iron Duke ?

6.—Name
Dogs that never bark.
A bird that is also a wine.
A fish that is also a weapon.
A flower that is also a musical instrument.
A horse that cannot be driven.

7.—Name the man who
Fiddled while Rome was burning.
Never told a lie.
Put the telescope to his blind eye.
Crossed the Rubicon.
Cut the Gordian knot.

8.—In what year did the following occur ?
Norman Conquest.
Gunpowder Plot.
Waterloo.
Execution of Charles I.
Outbreak of Boer War.

9.—What do the following stand for ?
C.O.D.
f.o.b.
C.U.A.C.
D.S.O.
O.M.

10.—In what plays do the following characters appear ?
Malvolio.
Dogberry.
Peter Quince.
Cassius.
Mr. Dangle.

11.—Name 5 cathedral towns in Yorkshire :—
1
2
3
4
5

12.—What is the cost of the following?
A letter of $2\frac{1}{2}$ ozs. to London.
A letter under 1 oz. to Australia.
A postcard to South Africa.
A postcard to France.
A parcel of 6 lbs. to Cornwall.

13.—Who invented :—
The Telephone.
The Phonograph.
The Steam Turbine.
The Thermionic Valve.
The X-ray Tube?

14.—In what countries are the following languages naturally spoken?
Erse.
Tamil.
Haussa.
Singhalese.
Provençal.

15.—Name a book of
Stevenson's without a heroine.
Thackeray's without a hero.
Dickens's without an end.
Scott's without a happy ending.
The Bible without a mention of God.

16.—Who said
Am I my brother's keeper?
Can there any good thing come out of Nazareth?
To obey is better than sacrifice.
Are not Abana and Pharpar better than all the waters of Israel?
My father chastised you with whips, but I will chastise you with scorpions.

17.—Give the feminine of the following :—
Fox.
Bachelor.
Czar.
Testator.
Colt.

18.—How would you begin a letter to the following ?
English Bishop.
An Archdeacon.
English Judge.
Duchess.
Dean.

19.—What professional man would be consulted if there had to be :—
A school planned.
A bridge built.
A will made.
A new area mapped.
An offending molar removed ?

20.—With what countries are the following men associated ?
Kossuth.
Rabindranath Tagore.
Prince Henry the Navigator.
Blasco Ibanez.
Capablanca.

21.—In what books or plays do the following appear ?
Mark Tapley.
Rochester.
Touchstone.
Tony Lumpkin.
Wendy.

22.—What Art was practised by the following ?
Pavlova.
Tagore.
Debussy.
Vermeer.
Rodin.

23.—What Science is connected with
Life.
The Mind.
God.
The Weather.
Microbes ?

24.—Which Book in the Bible tells of
The Prodigal Son.
A Slave's Return.
Four Horsemen.
A Shipwreck.
A Temperance Party?

25.—What is the name given to the young of the following animals?
Deer.
Eagle.
Swan.
Eel.
Goose.

ANALYSIS INTO SCORE GROUPS

Taking each Form List in turn, it is now possible to sort out the scores for each age into score groups. The range for each score group will depend on the circumstances. If the marks range from 0 to 100, it is convenient to take the score group from 0 to 9, 10 to 19, 20 to 29, etc., etc. All the boys whose scores fall between 40 and 49 are counted in that score group and, for all practical purposes, if the numbers are big enough, those scores will fall around a median of 45, since it is quite as likely for the same number to have scores from 45 to 49 as it is for others to have scores from 40 to 44.

Table X shows the distribution of scores in age groups for the above General Paper, and this table can be used to draw the histogram and the percentile curves for each age.

In drawing the percentile curves, the work can be conveniently arranged as follows. The graph paper is suitably marked in score groups, and the number of cases falling in each score group for each age is set out in columns as shown in Table XI (p. 77). Usually this can be done on the side of the graph

TABLE X

General Paper. March, 1928

	Age 18.	Age 17.	Age 16.	Age 15.	Age 14.
240					
230					
220					
210					
200					
190					
180				/ 1	
170				/ 1	
160		/ 1			
150	/ 1		/ 1		
140	/ 1	/ 1	/// 3	/ 1	/ 1
130		/// 3	// 2	// 2	// 2
120	/ 1	/ 1		/// 3	/ 1
110		///// 5	///// 5	/////// 7	// 2
100	/ 1	// 2	///// 5	////////// 11	/////// 7
90	/ 1	/ 1	////// 6	////// 6	/////// 7
80	/ 1	// 2	///// 5	////////////// 14	/////// 8
70		/ 1	//////// 8	////////////// 15	/////// 8
60		// 2	/ 1	//// 4	////////// 12
50				/////// 7	//// 5
40				//////// 9	////////////// 14
30				//// 4	//////// 9
20			/ 1	//// 4	/////// 7
10				/ 1	// 2
0					
	6	19	37	90	85

TABLE X (*continued*)

General Paper. March, 1928

	Age 13.	ge 12.	Age 11.	Age 10.	Age 9.	Age 8.
240						
230						
220						
210						
200						
190						
180						
170						
160						
150						
140						
130	/ 1					
120	/ 1	/ 1				
110	/ 1	/// 3		/ 1		
100	/ 1	/ 1				
90	/ 1	/ 1				
80	///// 5	//// 4		/ 1		
70	////////////// 13	// 2	/// 3			
60	////////////// 14	///// 5	/////// 7			
50	////////////////// 18	////// 6	////////// 11	/// 3	/ 1	/ 1
40	///////// 10	////////// 10	//////// 9	//// 4	//// 4	
30	//////////////////// 21	//////////////////// 20	////////// 11	//////// 8	///// 5	
20	//// 4	///// 5	////////////// 15	//////////////// 15	//// 4	/ 1
10	/// 3	//////// 8	///// 5	/////// 7	////////// 12	//////// 8
0		/ 1	/// 3	// 2	//// 4	////////// 10
	93	67	64	41	30	20

paper so that the data and the curves are shown on the same sheet. For clarity we have separated the Tables from the graphs.

Consider the 15-year-old Group (p. 77). Column I gives the frequency in each score group. Column II gives the integrated frequency.

Thus 1 boy scores between 10 and 19.
4 boys score „ 20 and 29.
Hence 5 „ „ below 30.
Also 4 „ „ between 30 and 39.

Hence 5 + 4 or 9 boys is the integrated frequency below 40.

So the integrated total is obtained by adding up all the scores to a given score group, as will be seen from Table XI. The total number of boys in this group is 90 and that is the integrated total at the top of Column II. These integrated totals are then reduced to a percentile basis and placed in Column III, where they now become percentiles.

This reduction is easily done by placing 90 over 100 on a slide-rule and by reading off the corresponding answers. From an inspection of curve 15 (Fig. 16) we see that the median score corresponding to the 50 percentile is 75, the Upper Quartile score corresponding to the 75 percentile is 95 and the Lower Quartile score corresponding to the 25 percentile is 54.

Hence the interquartile range is $95 - 54 = 41$.

$\therefore$ The semi-interquartile range is $\pm 20\frac{1}{2}$.

This is a measure of the mean deviation.

Fig. 16 shows the percentile curves for all the age-groups drawn on the same sheet, and the Table given below gives the median and quartiles for each age-group. It will be seen from Fig. 16 that the percentile curves lie one above the other, and it is obvious

TABLE XI

General Paper, 1928

	Age 17 and 18.			Age 16.			Age 15.			Age 14.			Age 13.		
	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
240															
230															
220															
210															
200															
190															
180							1	90	100						
170							1	89	99						
160	1	25	100				0	88	98						
150	1	24	96	1	37	100	0	88	98						
140	2	23	92	3	36	97	1	88	98	1	85	100			
130	3	21	84	2	33	89	2	87	97	2	84	99	1	93	100
120	2	18	72	0	31	84	3	85	94	1	82	96	1	92	99
110	5	16	64	0	31	84	7	82	91	2	81	95	1	91	98
100	3	11	44	5	31	84	11	75	83	7	79	93	1	90	97
90	2	8	32	5	26	70	6	64	71	7	72	85	1	89	96
80	3	6	24	6	21	57	14	58	64	8	65	77	5	88	95
70	1	3	12	5	15	40	15	44	49	8	57	67	13	83	89
60	2	2	8	8	10	27	4	29	32	12	49	58	14	70	75
50	0	0	0	1	2	5	7	25	28	5	37	43	18	56	60
40				0	1	3	9	18	20	14	32	37	10	38	41
30				0	1	3	4	9	10	9	18	21	21	28	30
20				1	1	3	4	5	6	7	9	10	4	7	7
10				0	0	0	1	1	1	2	2	2	3	3	3
0							0	0	0	0	0	0	0	0	0

Column I = Frequency.
Column II = Integrated Frequency.
Column III = Percentile.

TABLE XI (*continued*)

General Paper, 1928

	Age 12.			Age 11.			Age 10.			Age 9.			Age 8.		
	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
240															
230															
220															
210															
200															
190															
180															
170															
160															
150															
140															
130															
120	1	67	100												
110	3	66	99				1	41	100						
100	1	63	94				0	40	97						
90	1	62	92				0	40	97						
80	4	61	91				1	40	97						
70	2	57	85	3	64	100	0	39	95						
60	5	55	82	7	61	95	0	39	95						
50	6	50	75	11	54	84	3	39	95	1	30	100	1	20	100
40	10	44	66	9	43	67	4	36	88	4	29	97	0	19	95
30	20	34	51	11	34	53	8	32	78	5	25	83	0	19	95
20	5	14	21	15	23	36	15	24	58	4	20	67	1	19	95
10	8	9	13	5	8	12	7	9	22	12	16	53	8	18	90
0	1	1	1	3	3	5	2	2	5	4	4	13	10	10	50

Column I = Frequency.
Column II = Integrated Frequency.
Column III = Percentile.

that the General Paper was too hard for the 8-year-old group.

Age.	17.	16.	15.	14.	13.	12.	11.	10.	9.	8.
Median M. .	107	81	75	59	49	34	31	22	16	4
Q_1 . . .	127	100	95	81	65	56	48	32	27	7
Q_2 . . .	89	68	54	39	33	25	20	16	9	3
$Q=\frac{1}{2}(Q_1-Q_2)$	19	16	$20\frac{1}{2}$	21	16	$15\frac{1}{2}$	14	8	9	2
$V=\frac{100Q}{M}$.	18%	20%	27%	35%	32%	45%	45%	36%	56%	50%

It is sometimes useful to find out the average score in each age-group and to compare these with the median score. Various methods are given for finding the average scores in books on statistics, but results can probably be obtained just as quickly by means of a schedule drawn up as in Table XII (p. 81). Across the top of the schedule are the age-groups and running down the right-hand side are the Forms. From each Form List can be entered the number of boys in each age-group and their total score. When this is completed, the average is found by adding up in columns the total number of boys and their total scores.

This is illustrated in Table XII for the General Paper (p. 69) the graphical results for which are shown in Fig. 16. N represents the number of boys and S the total score.

For comparison purposes, the averages so obtained are tabulated with the corresponding medians obtained from the percentile curves.

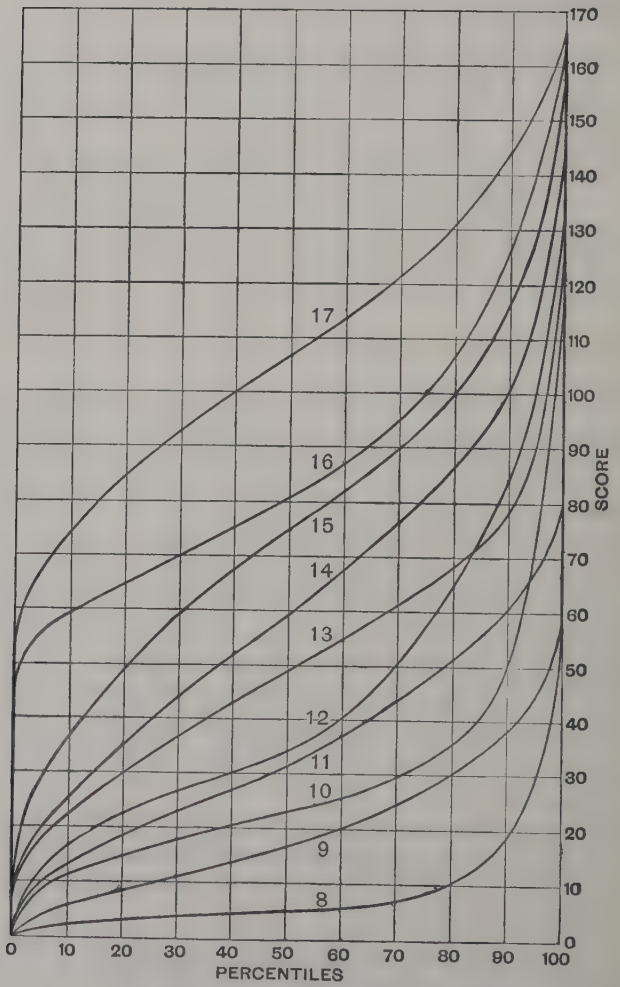

FIG. 16.—General Paper, April, 1928.

TABLE XII

AVERAGE SCORES AT EACH AGE

General Paper, April, 1928

Age 18		17		16		15		14		13		12		11		10		9		8		*Form*
N.	S.	N.	S.	N.	S.	N.	S.	N.	S.	N.	S.	N.	S.	N.	S.	N.	S.	N.	S.	N.	S	
1	144	3	444	3	294	9	1,063															A
1	95	2	168	5	422	7	681															B
2	224	6	685	2	245	2	241															C
2	237	4	484	6	766	7	682															D
6	700					9	967	12	1,102	3	287											E
				6	516	7	596	11	1,097			1	117									F
Av. 117		4	291	10	771	7	572	6	447													G
		19	2,072	2	156	17	1,276	4	359	1	49											H
				2	151	12	731	5	256													I
		Av. 109						8	423	17	1,352	6	483									J
						1	52	10	736	17	1,085											K
				1	22	8	266	11	488	5	250											L
				37	3,343					6	379	18	1,144	3	195							M
								4	106	13	503	5	123	3	125							N
				Av. 90·3		4	129	11	419	8	312	3	102									O
						90	7,256					7	355	17	928	3	192					P
								1	47	4	138	5	207	14	523	2	46					Q
						Av. 80·6		1	61	16	712	10	378	3	136							R
								1	17	3	82	12	275	7	165	2	75					S
								85	5,558	93	5,149	67	3,184	9	264	16	555	5	188			T
														7	141	11	325	3	65			U
								Av. 65·4		Av. 55·4		Av. 47·5				1	20	11	313	2	64	V
														1	22	6	82	7	88	3	18	W
														64	2,499	41	1,295	14	45	15	162	X
														Av. 39·0		Av. 31·6		40	699	20	244	24
																		Av. 17·5		Av. 12·2		Forms

Averages :	18	17	16	15	14	13	12	11	10	9	8
	117	109	90·3	80·6	65·4	55·4	47·5	39	31·6	17·5	12·2

Age	18.	17.	16.	15.	14.	13.
Average . .	—	109	90·3	80·6	65·4	55·4
Median . . .	—	107	81	75	59	49

Age	12.	11.	10.	9.	8.
Average . .	47·5	39	31·6	17·5	12·2
Median . . .	34	31	22	16	4

These results illustrate how a few good boys raise the average score above the median score. This is also evident from the rapid rise of the percentile curves after passing the median point. (Fig. 16.)

Example of Graphical Methods

Percentile and Histogram Distribution for an Intelligence Test. (*Fig.* 17.)

By casting the scores into score groups irrespective of age and arranging in columns, we obtain the following table:

Score Group.	Frequency.	Integrated Frequency.	Percentiles.
200–180 . .	2	394	100
179–160 . .	31	392	99
159–140 . .	84	361	92
139–120 . .	96	277	70
119–100 . .	101	181	46
99–80 . .	58	80	20
79–60 . .	19	22	6
59–40 . .	3	3	1

TO GET THE HISTOGRAM

Three boys gained scores between 40 and 59, and a rectangle of height 3 units was erected on the base

40–60. Nineteen boys gained scores between 60 and 79, and a rectangle of height 19 units was erected on the score base 60–80. Similarly, rectangles of heights 58, 101, 96, 84, 31 and 2 were erected opposite the corresponding score groups and the shaded histogram obtained is shown in Fig. 17.

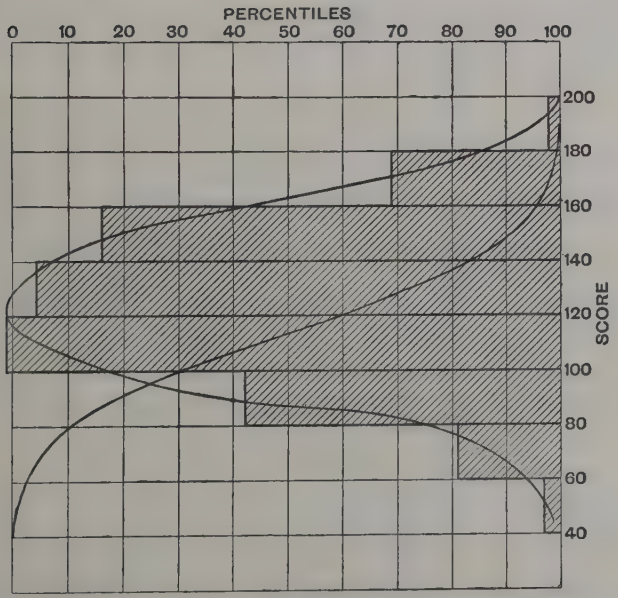

FIG. 17.—Group Test 33. Percentile and Histogram Distribution of 394 cases.

Score Groups.	Totals.	Sub-Totals.	Per-cents.
180	2	394	100
179–160	31	392	99
159–140	84	361	92
139–120	96	277	70
119–100	101	181	46
99– 80	58	80	20
79– 60	19	22	6
59– 40	3	3	1

A smooth line, called the curve of "closest fit", was then drawn so as to include roughly as much area as the histogram. This curve is the frequency curve and from it can be read off the probable frequency of any given score. The curve in Fig. 17 is reasonably symmetrical, thus showing that the distribution of intelligence was approximately normal. This, of course, is also shown by the fact that the totals in Column I rise from 2 to 101 and then fall off again to 3.

On the same graph the percentile distribution irrespective of age was obtained as explained above (p. 28). In the second column, the integrated total is shown, and this integrated total is reduced to a percentile basis. The percentiles are then set off opposite the *middle* of the corresponding score group and a smooth percentile curve drawn. From this it will be seen that the median score, or the 50 percentile, is 114, the Upper Quartile, or 75 percentile, is 133, and the Lower Quartile, or the 25 percentile, is 93.

Hence, half the boys obtained scores between 93 and 133.

Also quarter of the boys obtained scores above 133.

And „ „ „ „ „ below 93.

The range 133 minus 93 equals 40, which is the measure of the dispersion of the scores.

PERCENTILE CURVES FOR EACH AGE-GROUP

In the above, all the boys have been treated as one group irrespective of their ages. Then the scores were cast into age-groups and the following Table gives the data obtained for three age-groups.

Score Groups.	Age 16.			Age 15.			Age 14.		
	I	II	III	I	II	III	I	II	III
200–180				1	83	100			
179–160	8	66	100	9	82	99	3	104	100
159–140	21	58	88	15	73	88	20	101	97
139–120	17	37	56	26	58	70	24	81	78
119–100	14	20	30	19	32	39	23	57	55
99– 80	6	6	9	9	13	16	23	34	33
79– 60				3	4	5	10	11	11
59– 40				1	1	1	1	1	1

Column I = Frequency.
Column II = Integrated Frequency.
Column III = Percentile.

The corresponding percentile curves were drawn, as shown in Fig. 18. From these curves we find the following results:

	Age 16.	Age 15.	Age 14.
Median M	124	118	103
Q_1	142	138	125
Q_2	105	98	81
$Q = \frac{1}{2}(Q_1 - Q_2)$ . .	$18\frac{1}{2}$	20	22
$V = \frac{100Q}{M}$	15%	17%	21%

COMPARISON OF FORMS

It is often useful to compare the results obtained in different Forms, and a handy method of doing this is illustrated in Fig. 19. The scores are set off on the left-hand side of the graph and the Forms starting from the top one are set off at equal intervals at the bottom of the graph. For each Form a line is drawn between the Upper limit and the Lower limit, and

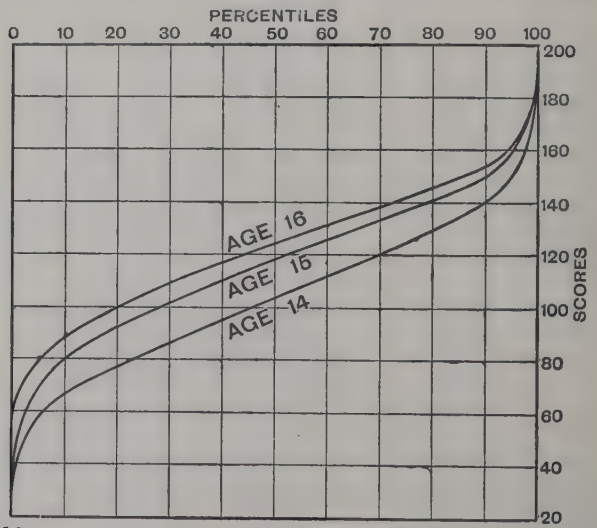

FIG. 18.—Group Test 33. Percentile distributions. Ages 14–15–16

on this line the positions of the Upper Quartile, Median, the Mean or average score and the Lower Quartile can be marked. Fig. 19 illustrates the results of an Intelligence Test given to 14 Forms. One can make deductions about the ability of each Form by inspecting the Figure. If the median is more than half-way up, it means that the top half of the Form is good. If the Lower Quartile and the

Lower Limit are low, it means that the bottom quarter of the Form is poor.

The comparison of Form by Form is also easily seen by the dotted lines which join the quartiles, medians and the averages. If the average is below the median, it is probably due to the fact that a boy

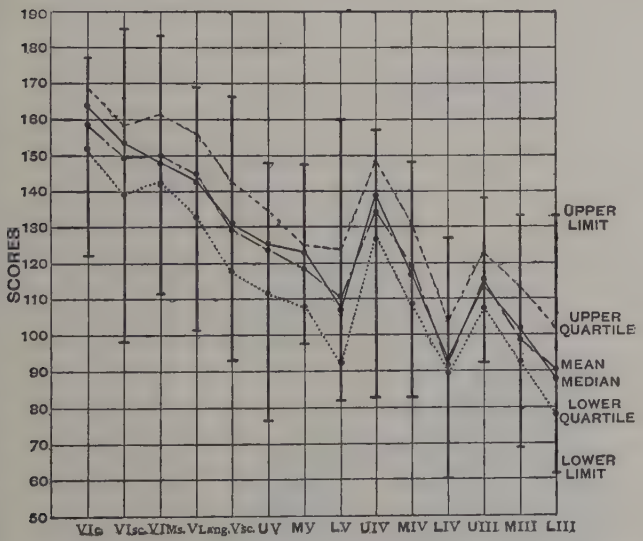

FIG. 19.—Group Test 33. Variations in the Tests from Form to Form.

with a poor score has depressed the average score. On the other hand, a boy with a very good score may raise the average above the median for a Form. For this reason, it is probably better to take the median than the average, because the median is not so influenced by very good or very bad scores.

Latin Grammar Paper (March, 1927).

The following Latin Grammar Paper was given to all boys doing Latin and the Histogram and Percentile Curves drawn.

Name.....

Form.....

Age on July 31st last.....

LATIN ACCIDENCE PAPER

Time—1 *Hour.*

All answers are to be written opposite to the questions on the RIGHT.

100 *Marks.*

1.—	Genitive Singular.	insula.
	Dative Singular.	vir.
	Dative Plural.	ager.
	Nominative Plural.	regnum.
	Accusative Plural.	deus.
2.—	Genitive Plural.	tempus.
	Ablative Singular.	genus.
	Dative Singular.	exercitus.
	Nominative Plural.	genu.
	Genitive Plural.	hostis.
3.—	Ablative Singular.	ovile.
	Accusative Plural.	domus.
	Vocative Singular.	gradus.
	Accusative Singular.	nomen.
	Genitive Plural.	dux.
4.—	Dative Plural.	filia.
	Dative Singular.	res publica.
	Genitive Singular.	volgus.
	Ablative Plural.	vis.
	Genitive Plural.	iter.

5.—Ablative Fem. Singular. miser.
Ablative Fem. Singular. celer.
Ablative Fem. Singular. sacer.
Ablative Fem. Singular. facilis.
Ablative Fem. Plural. pius.

6.—Comparative. felix.
Comparative. magnus.
Comparative. utilis.
Superlative. malus.
Superlative. difficilis.

7.—eighteen.
eighty.
eight thousand.
eleventh.
four hundredth.

8.—Genitive Singular. tu.
Dative. se.
Dative Singular. ille.
Genitive Singular. alius.
Genitive Plural Masc. idem.

9.—with whom (plural).
each of two (feminine).
four times.
at Eboracum.
two camps.

10.—3rd plur. fut. indic. sum.
2nd sing. perf. indic. act. laudo.
1st sing. fut. indic. act. audio.
2nd plur. pluperf. indic. act. rego.
3rd sing. present indic. act. moneo.

11.—1st plur. perf. subj. act. rego.
2nd plur. imperf. subj. act. audio.
3rd plur. pluperf. subj. act. moneo.
1st sing. pres. subj. act. laudo.
3rd sing. imperf. subj. act. moneo.

12.—2nd sing. imperf. indic. pass. rego.
3rd sing. perf. indic. pass. moneo.
3rd plur. fut. perf. indic. pass. rego.
2nd plur. pres. indic. pass. audio.
1st plur. fut. indic. pass. laudo.

13.—1st plur. pres. subj. pass. laudo.
2nd sing. imperf. subj. pass. moneo.
3rd plur. perf. subj. pass. rego.
1st sing. pluperf. subj. pass. audio.
2nd plur. imperf. subj. pass. audio.

14.—3rd sing. imperative act. moneo.
perfect infin. act. rego.
genitive gerund. audio.
future participle masc. laudo.
present participle. audio.

15.—perf. participle pass. (feminine). moneo.
gerundive (masculine). rego.
present infin. pass. rego.
perfect infin. pass. audio.
future infin. pass. laudo.

16.—3rd sing. pres. indic. volo.
3rd plur. perf. indic. nolo.
2nd sing. pres. indic. act. fero.
1st plur. imperf. indic. possum.
1st sing. fut. indic. eo.

17.—he remembered.
says he.
thou art being carried (use fero).
to prefer.
2nd plural imperf. subj. fio.

18.—1st sing. perf. indic. act. mitto.
1st sing. perf. indic. act. aperio.
1st sing. perf. indic. act. pello.
Supine. rumpo.
Supine. quaero.

19.—1st sing. perf. indic. sequor.
1st sing. perf. indic. orior.
Supine. fingo.
Supine. vincio.
Present Infin. proficisor.

20.—Without Cæsar.
Before Cæsar.
Around Cæsar.
Outside Cæsar.
In the power of Cæsar.

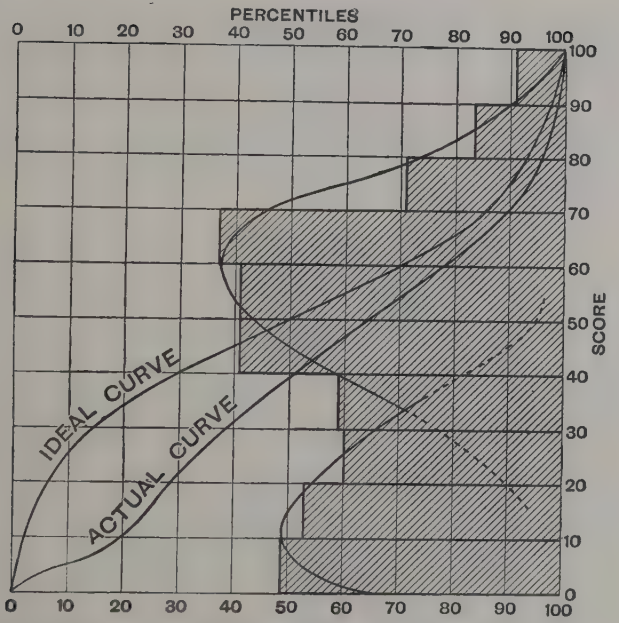

FIG. 20.—Latin Grammar Paper, March, 1927. 413 cases.

Theoretical Median 50 (Curve drawn from Normal Curve on 5σ Base)
Actual ,, 40
64 per cent of Boys below Half Marks.

Score Group.	COLUMN I. Frequency.	COLUMN II. Integrated Frequency.	COLUMN III. Percentiles.
100–90	8	413	100
89–80	16	405	98
79–70	29	389	94
69–60	63	360	87
59–50	59	297	72
49–40	59	238	58
39–30	41	179	43
29–20	40	138	34
19–10	47	98	24
9– 0	51	51	12

Fig. 20 shows the distribution of scores irrespective of the age. It is obvious that the histogram is bimodal and that probably there were two distinct groups of candidates. Hence, it would be better in future to set two papers—one higher and one lower —for one paper was evidently not suitable for the great range covered.

The percentile curve shows that the median score, or 50 percentile, was 40 and that 64 per cent of the boys were below half marks. It is seen that the experimental curve falls below the percentile curve that would be found for an ideal and normal distribution of scores. Hence, either the Latin Grammar was weaker than it should be, or the paper was too hard for the majority of the boys. The fact that the histogram is bimodal shows that the second supposition is more likely. Too many boys got low marks because of the fact that they had only been doing Latin a short time and they were only able to do accurately a portion of the paper.

The scores were then arranged in score groups for each age and the percentile curve for each age drawn (Fig. 21).

From the curves the medians and quartiles can be found and hence the Quartile Deviation, which is a measure of the dispersion of the scores in each age-group.

It will be seen that the 15-year-old group was most homogeneous and that the poorer 15-year-old boys were better than the poorer 16-year-old boys, though the good 16-year-old boys beat the good 15-year-old boys. Thus, so far as Latin Grammar was concerned, the 16-year-olds had a rather poor tail. In the following year the results confirmed the suspicion aroused by the above tests.

As will be seen from the results (p. 94), the histogram is again bimodal, showing the existence of two distinct

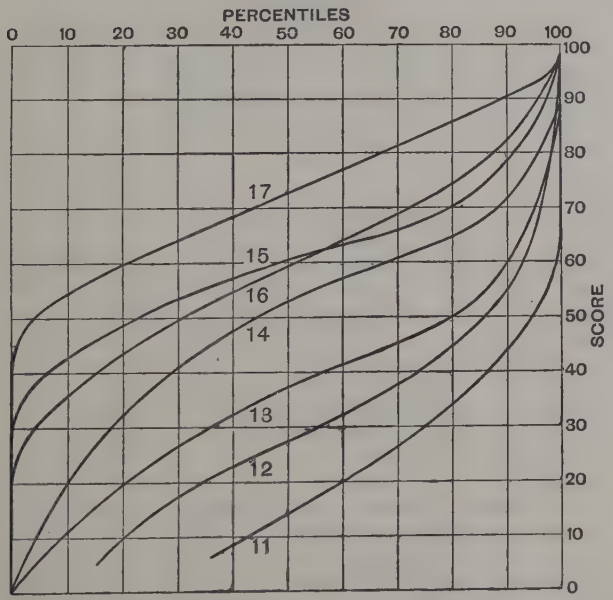

FIG. 21.—Latin Grammar Paper, March, 1927. Ages 17–11.

RESULTS

Age . .	17	16	15	14	13	12	11
Median .	73	59	60	52	37	28	14
Q_1 . .	84	71	69	63	48	40	30
Q_2 . .	62	48	51	37	23	13	—
$\frac{1}{2}(Q_1-Q_2)$	11	$11\frac{1}{2}$	9	13	$12\frac{1}{2}$	$13\frac{1}{2}$	15
V . .	15%	20%	15%	25%	34%	48%	100%

groups. The percentile curves show that the relationship between the 15-year-olds and the 16-year-olds is roughly the same as in the previous year.

LATIN GRAMMAR PAPER, 1928

Histogram of 459 Cases

All Ages 18–9

Score Groups.	Column I. Frequency.	Column II. Integrated Frequency.	Column III. Percentiles.
100 . . .	2	459	100
99–90 . . .	13	457	100
89–80 . . .	36	444	97
79–70 . . .	68	408	89
69–60 . . .	69	340	74
59–50 . . .	59	271	59
49–40 . . .	43	212	46
39–30 . . .	40	169	37
29–20 . . .	49	129	28
19–10 . . .	54	80	17
9–0 . . .	26	26	6

Results obtained from Percentile Curves.

Age . .	17	16	15	14	13	12	11
Median .	55	67	67	59	49	30	22
Q_1 . .	72	77	76	69	62	50	37
Q_2 . .	36	54	58	44	32	16	10
$\frac{1}{2}(Q_1-Q_2)$	18	$11\frac{1}{2}$	9	$12\frac{1}{2}$	15	17	$13\frac{1}{2}$
V . .	33%	17%	13%	21%	30%	57%	61%

It should be noted that these 15-year-olds were the previous year's 14-year-olds, and that these 16-year-olds were the 15-year-olds of the previous year, who at that time were better on the whole than boys a year older. This extra improvement of the 15-year-olds compared with the age group in the year above may be explained in various ways.

(1) There might be a particularly good teacher exerting his influence at this point in the School.

(2) The brighter 15-year-olds might have left and so were not in the 16-year-old group in the next year.

(3) The 15-year-olds were mostly working for School Certificate and were at the top of their form in Latin, while many of the 16-year-olds, on passing into the Sixth Form, had specialized and dropped Latin as a leading subject.

(4) The type of paper might suit the 15-year-old candidates better than the others.

Some Latin Sentences were given to the Fifth Forms and the results graphed.

RESULTS

Score Group.	Column I. Frequency.	Column II. Integrated Frequency.	Column III. Percentiles.
100–90 . . .	2	78	100
89–80 . . .	11	76	97
79–70 . . .	27	65	83
69–60 . . .	18	38	49
59–50 . . .	8	20	25
49–40 . . .	11	12	15
39–30 . . .	1	1	1
29–20 . . .	0	0	0

The percentile curve was reasonably normal and gave a median score of 65.

Also $Q_1 = 72$

$Q_2 = 55$

Hence $Q = \frac{1}{2}(Q_1 - Q_2) = \pm 8{\cdot}5$

and $V = \frac{100Q}{M} = 13{\cdot}1$ per cent

The fact that the Mean Deviation was $\pm$ 8·5 and the Variability 13·1 per cent showed that the work was quite satisfactorily homogeneous.

Name.....
Form.....
Age last July.....

ACCURACY PAPER

TIME ALLOWED.—ONE HOUR

Work out the following questions in any order.
Give your answers on right-hand side of each question.
Give answers in units shown on right-hand side.
Hand in both question paper and working paper.
All tables required are at the bottom of the paper.

4 marks per question. Total 100.

1. 537 + 1428 + 32 + 1875
2. 5 hrs. 23 mins. 55 secs. + 6 hrs. 12 mins. 37 secs.
3. 387623 — 128579
4. 45 yds. 1 ft. 8 ins. — 17 yds. 2 ft. 10 ins.
5. 987 × 789
6. £4 7s. 2½d. × 14
7. 44232 ÷ 97
8. 40 tons 19 cwts. 2 qrs. ÷ 11
9. ·8725635—·868479
10. 5·76 + ·1234 — ·0726 — ·9874
11. ·183924 ÷ ·234

ANSWERS.

1. .....
2. .....hrs. .....mins. .....secs.
3. .....
4. .....yds. .....ft. .....ins.
5. .....
6. £..... .....s. .....d.
7. .....
8. .....tons. .....cwts. .....qrs.
9. .....
10. .....
11. .....

12.	$\frac{\quad}{\cdot 273}$	12. .....
13.	Reduce $\frac{6912}{7344}$ to its lowest terms.	13. .....
14.	$3\frac{5}{8} - 4\frac{3}{4} + 7\frac{7}{16}$	14. .....
15.	$\frac{56}{65} \times \frac{45}{77} \times \frac{91}{48}$	15. .....
16.	$\frac{\frac{5}{8} \times \frac{12}{25}}{\frac{14}{121} \times \frac{11}{7}}$	16. .....
17.	Reduce 15 qrs. 17 lbs. 11 ozs. to ounces.	17. .....ozs.
18.	Reduce 5 sq. yds. 7 sq. ft. 32 sq. ins. to sq. ins.	18. .....sq. ins.
19.	4 sq. yds. 8 sq. ft. 79 sq. ins.—2 sq. yds. 7 sq. ft. 93 sq. ins.	19. .....sq. yds.....sq. ft.....sq. ins.
20.	102 sq. yds. 12 sq. ft. 135 sq. ins. ÷ 15	20. .....sq. yds.....sq. ft.....sq. ins.
21.	Reduce 8 cu. yds. 23 cu. ft. 1243 cu. ins. to cu. ins.	21. .....cu. ins.
22.	$\frac{\text{3 cu. ft. 382 cu. ins.}}{\text{1 ft. 10 ins.}}$	22. .....sq. ft. .....sq. ins.
23.	Find the weight of water in a cistern 7′ 9″ by 5′ 9″ by 4′ 6″	23. .....qrs. .....lbs. .....ozs.
24.	Express £4 16s. 10d. as a percentage of £5 16s. 8d.	24. .....%
25.	A form of 20 boys has an average age 12 yrs. 8 mths. What will be the new average age (to the nearest month) if 4 boys, each 14 yrs. 1 mth. old, are added to the form ?	25. .....yrs.....mths.

1 hr. =60 mins. 1 yd.=3 ft. 1 ton =20 cwts. 1 sq. yd.=9 sq. ft. 1 cu. yd.=27 cu. ft.
1 min.=60 secs. 1 ft. =12 ins. 1 cwt.=4 qrs. 1 sq. ft. =144 sq. ins. 1 cu. ft. =1728 cu. ins.
Wt. of 1 cu. ft. of water=62½ lbs. 1 qr.=28 lbs. 1 lb.=16 ozs.
Average age= $\frac{\text{Total age of Form}}{\text{Total number of boys.}}$

$1\% = \frac{1}{100}$

Accuracy Paper

An Accuracy Paper was given to a number of Forms and the percentile curves drawn. The values of the Medians and the Quartiles are given below.

It will be seen that the percentage variation from the mean was greatest in the 12-year-old group and least in the 16-year-old group.

Age . . .	17	16	15	14	13	12	11	10
Median M.	46	45	41	38	27	23	23	13
Q_1 . .	57	52	51	51	39	35	31	18
Q_2 . .	35	36	28	25	17	15	14	9
$\frac{1}{2}(Q_1-Q_2) = Q$	± 11	± 8	$\pm 11\frac{1}{2}$	± 13	± 11	± 10	$\pm 8\frac{1}{2}$	$\pm 4\frac{1}{2}$
$V = \frac{100Q}{M}$	24%	18%	28%	34%	41%	43%	37%	35%

The marks were low because the paper was rather hard and the time allowed was on the short side. The following year an easier paper was set. This is given on page 96.

The results obtained from drawing the percentile curves were as follows :

Age . . .	17	16	15	14	13	12	11	10
Median .	73	69	62	57	56	43	36	28
Q_1 . .	82	80	74	71	69	52	49	36
Q_2 . .	55	58	49	42	42	31	24	22
Q . . .	$13\frac{1}{2}$	11	$12\frac{1}{2}$	$14\frac{1}{2}$	$13\frac{1}{2}$	$10\frac{1}{2}$	$12\frac{1}{2}$	14
V . . .	$18\frac{1}{2}$%	16%	20%	25%	24%	$24\frac{1}{2}$%	35%	50%

These results show that—

(1) The 16-year-old group was the most homogeneous.

(2) The 13-year-old group was practically as good as the 14-year-old group.

French Grammar Paper

A French Grammar Paper was set to eight Forms and the Scores were analysed into Score Groups for each age. The few 17-year-old boys were counted in with the 16-year-olds. The percentile curves were drawn and an analysis showed the following results :

Age	16	15	14	13
Median M . .	42	38	32	19
Q_1	53	51	48	29
Q_2	31	24	17	11
Q	11	$13\frac{1}{2}$	$15\frac{1}{2}$	9
V	21%	26%	48%	47%

The paper is given below.

Name..... *Age on July* 31*st*, 1928.....

Form..... *Time*, 1 *hour*.

FRENCH PAPER.

All answers to be written on this Paper opposite the question.

100 *marks*.

1.—Put into the plural :

1. le travail.
2. le bal.
3. l'œil.
4. le clou.
5. le chou.
6. le monsieur.
7. le grand-père.
8. l'arc-en-ciel.
9. égal.
10. bleu.

2.—Give the feminine of :

1. le maître.
2. le dieu.
3. le porteur.
4. un acteur.
5. le héros.
6. le neveu.
7. le mari.
8. le gendre.
9. le pécheur.
10. le compagnon.
11. fier.
12. mignon.
13. muet.
14. jumeau.
15. vieux.

3.—Put into French :

1. With sugar.
2. To friends.
3. Without money.
4. What a story.
5. As a soldier.
6. Per day.
7. Covered with snow.
8. Some fine horses.
9. To such women.
10. Many pears.
11. The majority of men.
12. A castle tower.
13. He has no books.
14. At mid-June.
15. Five francs a pound.

4.—What preposition is omitted ?

1. Il pria le juge ——— lui pardonner.
2. ——— avoir parlé, il s'assit.
3. Il finit ——— pleurer.
4. ——— faisant cela, vous réussirez.
5. Vous êtes trop jeune ——— savoir.

5.—Form adverbs from the following adjectives:

1. naturel.
2. mou.
3. prudent.
4. impuni.
5. bon.
6. fou.
7. exprès.
8. mauvais.
9. doux.
10. gentil.

6.—Give the comparative degree of:

1. mal.
2. petit.
3. bon.
4. beaucoup.
5. beau.

7.—Change the words in italics into pronouns, and make any other changes you consider necessary.

1. Jean est allé à *Jeanne.*
2. Donnez-moi *des joujoux.*
3. Il m'a présenté à *la dame.*
4. Il m'a présenté *sa sœur.*
5. Elles sont plus grandes que *leurs frères.*

8.—Put the verbs in italics into the correct tense.

1. Qui vivra, *voir.*
2. Il a peur qu'il ne *venir.*
3. S'il vient et qu'il vous *voir,* il nous parlera.
4. Il y a longtemps que je *être* ici.
5. S'il *découvrir* cela, vous serez perdu.

9.—What pronouns would complete the sense of the following :—

1. Il ne sait ——— faire.
2. Dites-moi ce ——— vous parlez.
3. ——— des poires est la meilleure ?
4. De ——— êtes-vous le fils ?
5. Tels étaient les hommes parmi ——— il s'était établi.
6. Voici l'homme à la bonté ——— je dois mon bonheur.
7. Notre maison n'est pas si grande que ——.
8. De ces deux livres je préfère ——— de mon frère.
9. C'est tout ——— je veux dire.
10. Il loue les serviteurs ——— choisit le roi.

10.—

1. 3rd sing., pres. indic., dormir.
2. 3rd plur., pres. indic., pouvoir.
3. 1st plur., imperf. subj., manger.
4. 2nd plur., future indic., lever.
5. 3rd plur., future indic., savoir.
6. 3rd plur., pres. indic., appeler.
7. 2nd sing., fut. indic., faire.
8. 2nd plur., imperf., indic., lier.
9. 1st sing., pres. subj., vouloir.
10. 1st plur., past perf., indic., devenir.
11. 1st plur., perf. indic., se nuire.
12. 3rd sing., fut. perfect, indic., devoir.
13. 3rd sing., pres. subj., aller.
14. 3rd sing., imperf. subj., tenir.
15. 3rd plur., perfect subj., arriver.
16. 2nd sing., imperative, aller.
17. 3rd plur., pres. indic., connaître.
18. 1st plur., fut. indic., courir.
19. 1st plur., pres. subj., craindre.
20. 3rd plur., past historic, indic., écrire.

THE USE OF PERCENTILE RANKS

These can often be used to compare any given boy with his fellows of the same age. Thus, if a boy undergoes a series of tests in Intelligence, General Knowledge, Accuracy, etc., from his scores his percentile position with regard to boys of his own age can be found immediately from the percentile curves which have been constructed for each age group.

Some extracts from typical record cards are shown, and it will be noticed that A is uniformly good and C is uniformly weak. B is a case of a boy just above the average except that he has done well in the Accuracy Papers. Considering that different papers were given each year to a large number of boys, the agreement amongst a boy's percentile ranks is very striking.

PERCENTILE RECORD CARD

Name.	Year.	Age.	Int. Test.	Gen. Paper.	Acc. Paper.	Lat. Gr.
A . .	1926–27	13·0	90	97	—	97
A . .	1927–28	13·7	90	99	98	99
A . .	1928–29	14·7	94	99	99	—
B . .	1926–27	12·11	65	79	—	62
B . .	1927–28	13·6	65	85	90	65
B . .	1928–29	14·6	66	—	94	—
C . .	1926–27	14·7	7	5	—	1
C . .	1927–28	15·2	4	10	—	1
C . .	1928–29	16·2	—	—	14	—

CHAPTER VI

SOME FURTHER APPLICATIONS

ALLOWANCE FOR AGE

IN a Scholarship Examination, where candidates may range between 10 and 12 years of age, it is usual to make some allowance to the younger candidates. On what basis should this be given? One method is to increase the marks of the candidates by a fixed percentage for each month that they are below 12 years of age. The difficulty is to settle what this monthly percentage allowance should be. Most schools in electing scholars are on the lookout for general ability and are out to defeat cramming. For this reason, some light may be thrown on the question of the monthly percentage allowance by a study of the norms for tests such as General Tests rather than the norms for Standard Tests in the regular school subjects, such as Arithmetic. In order to investigate this question, the median scores obtained from the percentile curves were plotted against the ages for three separate General Papers set to the whole School in 1927, 1928 and 1929. It was found that a straight line could reasonably be drawn through these median scores in each case.

The slopes of the straight lines were not the same, nor were the lines coincident. This is because the actual marks obtained depended on the difficulty of

the papers. The important point to investigate is how the norm changes in the neighbourhood of the norm for the 12-age group in each case. Since the lines were straight, the increase in the norm per year was constant for a given test, and the method adopted was to scale the norms so that the norm for 12-age group ranked as 100 and then the norms for boys of 10 and 11 were expressed on this scale.

The median scores to the nearest whole number read from the lines were as follows:

Norm	12	11	10
General Paper, 1927 . . .	71	56	41
General Paper, 1928 . . .	53	40	28
General Paper, 1929 . . .	45	34	22

Scaled in terms of the 12 norm being taken as 100 we get

Norm	12	11	10
General Paper, 1927 . . .	100	79	58
General Paper, 1928 . . .	100	76	53
General Paper, 1929 . . .	100	76	49
Average	100	77	53

Therefore the drop per annum is roughly 24. Hence a reasonable age scale would be

12 years of age, 100
11 ,, ,, ,, 76
10 ,, ,, ,, 52

The matter was tested in another way. The median scores at each age were added together and

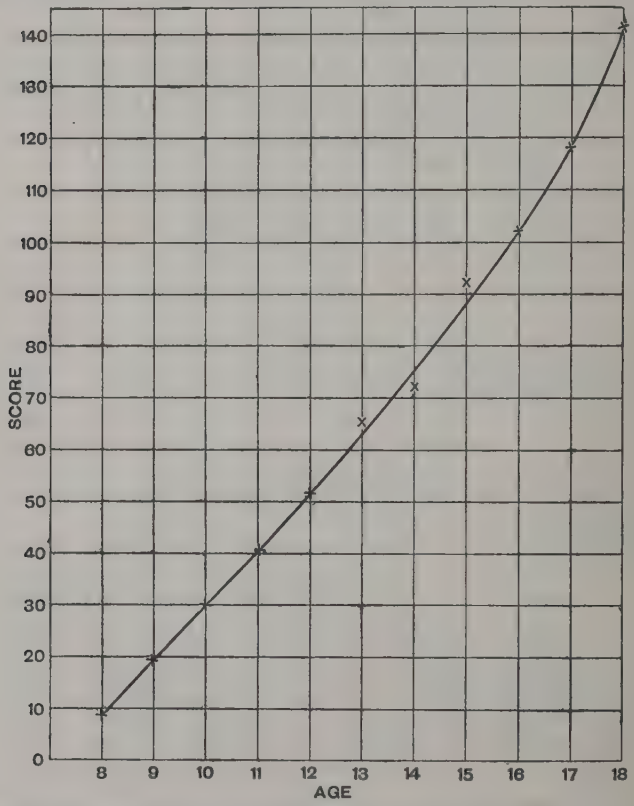

FIG. 22.—Graph of the average of the Median Scores in three General Papers, 1927–29.

Age . . .	8	9	10	11	12	13	14	15	16	17	18
Average Score over Three Tests . .	8	19	30	40	52	65	73	92	102	118	141

the average found and then plotted (Fig. 22). The values found were :

Age	18	17	16	15	14
Average Score over the 3 Tests	141	118	102	92	78

Age	13	12	11	10	9	8
Average Score over the 3 Tests	65	52	40	30	19	8

The graph is practically a straight line between ages 8 and 14 and then begins to curve upwards.

On the relevant portion between 8 and 13 the slope is about 11 or 12 marks per year. Thus, if the fixed point at 12 years of age be taken as 100 instead of 52, we get the scale scores as follows :

13	12	11	10
124	100	76	52

This then gives us the opportunity of constructing an age scale with the corresponding per cent mark to be expected if the boy is quite normal. Since the percentage mark drops by approximately 24 per year of age, we can assume as a rough working rule that the age allowance is 2 per cent per month. The proper adjustment would not be got by adding 2 per cent per month to a boy's actual marks. This would make some of the allowances too small. To take an example :

Boy.	Age.	Marks.	Months below 12.	Per cent Allowance.	Allowance.	New Total.	Rank.
A	11·10	591	2	4	24	615	2
B	11·9	564	3	6	34	598	4
C	11·3	532	9	18	96	628	1
D	11·4	526	8	16	84	610	3

Now we have shown that the proportion amongst the marks is Age 12 : 11 : 10
Mark 100 : 76 : 52.

Or in a Table:

I. Age.	II. Per cent.	III. Scale on Slide-Rule.	IV. Actual Marks.	V.	VI. Rank.
12·0	100				
11·11	98				
11·10	96	591	A 591	A is the fixed point	3
11·9	94	580	B 564	B is 16 marks too few	4
11·8	92	567			
11·7	90	555			
11·6	88	543			
11·5	86	530			
11·4	84	517	D 526	D is 9 marks in hand	2
11·3	82	505	C 532	C is 27 marks in hand	1
11·2	80	494			
11·1	78	481			
11·0	76	474			

In the above Table, taking the marks as 100 at 12, the marks that should be obtained by a boy compared with this fixed standard are shown for each month in Column II. The boy A with the most marks in the examination is age 11.10 and has got

591 marks. If we fix him as the standard, so that 591 corresponds to 96 at age 11.10, we get 580 corresponding to 94 at age 11.9, and 567 corresponding to 92 at age 11.8 by the slide-rule.

Hence, Column III gives the marks the others should obtain if they are as good as A, making allowance for their ages.

Column IV shows the actual marks obtained by A, B, C and D, and these marks are placed in their proper position in the age scale, A at 11.10, B at 11.9, C at 11.3 and D at 11.4.

Column V shows the deductions to be made. Compared with A at 591, B should have got 580 marks. He obtained 564, so that he is 16 marks below his true position. Compared with A at 591, C should have got 505. He obtained 532, so that he is 27 marks in hand. Compared with A at 591, D should have got 517. He obtained 526, so he is 9 marks in hand. Hence, the actual order is C, D, A, B, as is shown in Column VI.

Compared with the rough method of adding 2 per cent worked out above, this method, which is more accurate and fair, interchanges D and A in the final order.

The matter may be looked at in another way. A is the fixed standard. B gets 564, when he should have got 580. Hence B is one month below his standard position. C gets 532, which is the score corresponding to age 11.5. Hence C is two months above his standard position. D gets 526, which makes him almost a month above his standard position.

Hence, allowing for their ages, the order of the boys is First C, Second D, Third A, Fourth B, which is the same as the final result above.

It is not pretended that this method disposes of the question, but it is put on record as an attempt to place the age allowance on a proper basis. It brings out the point that a percentage allowance is not quite enough, since this penalizes the younger boys. The only fair method is to get out an ideal-mark standard for each age and then compare each boy's actual position with his standard position. The standard position will, of course, depend on further research into the norms for boys of different ages. It is not suggested that the above scale is the true one, but it is put forward as an empirical scale based on actual experimental results.

SCHOOL ESTIMATES AND EXAMINATION RESULTS

An interesting use of correlation is to compare the results of an external examination like the School Certificate with the estimate given by the masters actually teaching the boys. One method of doing this is to get every month an estimate of the percentage mark for every boy in every subject and then, based on the year's work, to get a final estimate before the examination. The master makes a return showing the actual percentage mark the boy is estimated to be worth and his rank in the Form. The School Certificate results give the marks for each boy in each subject, and from this an order for each subject for each Form can be obtained. The two sets of results can now be compared in two ways.

(1) *By the Method of Ranks.*

The correlation coefficient between the master's estimate and the School Certificate result can be worked out in each subject and tabulated. Some

results for four years are shown below. Care must be made in making deductions, but a low correlation coefficient may be due to—

1. The examination result being at fault.
2. The masters' estimate being at fault.
3. The master and the examining body not being in agreement on the outstanding features of the syllabus.

For a given subject, say, Science, the estimates are given by the several masters who are taking the Forms, and if the Science correlations are all low, it suggests that the technique of examining in Science is not so well developed as in other subjects, or that Science is a difficult subject to examine.

Low correlation in the weaker forms seems to point to the fact that examination results are not too trustworthy for border-line cases. If a Form is not homogeneous, it ought theoretically to be easier to get agreement in the results. Experience at present does not seem to bear this out, for instances crop up of extraordinary examination results. The results to date show that it is generally easier to get high correlations with the good forms and that it is easier in some subjects than in others. The high correlation in most cases shows that the masters' estimates and the examination results are remarkably trustworthy.

CORRELATION COEFFICIENTS, 1926

	A.	B.	C.	D.
English	0·72	0·51	0·36	0·41
History	0·80	0·61	0·67	0·75
French	0·83	0·87	0·44	0·53
Latin	0·81	0·74	0·55	—
German	0·59	—	—	—
Greek	0·79	—	—	—
Mathematics	0·79	0·51	0·65	0·45
Physics	—	0·51	0·70	—
Chemistry	0·69	0·50	0·25	—
Physics and Chemistry	—	—	0·50	0·61
Geography	—	—	0·26	0·56
Mechanics	—	0·88	—	—

CORRELATION COEFFICIENTS, 1927

	A.	B.	C.	D.	German Set.	Mechanics Set.
English	0·61	0·67	0·37	0·46		
History	0·75	0·73	0·73	0·69		
French	0·63	0·77	0·82	0·68	0·84	0·22
Latin	0·77	0·73	0·79	0·40		
Greek	0·67	—	—	—		
Mathematics	0·70	0·79	0·75	0·80		
Physics	0·23	0·74	0·78	0·32		
Chemistry	0·76	0·84	0·58	0·18		
Geography	—	0·78	0·41	—		

CORRELATION COEFFICIENTS, 1928

	A.	B.	C.	D.	Sets.	
English . .	0·55	0·56	0·44	0·57	Greek	0·85
History . .	0·87	0·69	0·77	0·84	Mechanics	0·50
French . .	0·64	0·64	0·65	0·51	Geography	0·37
Latin . .	0·78	0·87	0·62	0·79	Biology	0·45
Mathematics	0·74	0·89	0·70	0·83	German	0·89
Physics . .	0·76	0·81	0·68	0·62		
Chemistry .	0·76	0·65	0·83	0·79		

CORRELATION COEFFICIENTS, 1929

	A.	B.	C.	D.	Sets.	
English . .	0·73	0·61	0·63	—	Greek	0·83
History . .	0·71	0·61	0·89	0·63	Mechanics	0·84
French . .	0·71	0·82	0·55	0·44	Geography	0·74
Latin . .	0·75	0·66	0·65	0·16	Biology	0·53
Mathematics	0·75	0·83	0·89	0·53	German	0·88
Physics . .	0·68	0·76	0·60	0·58		
Chemistry .	0·80	0·70	0·68	0·83		

AVERAGES OF CORRELATION COEFFICIENTS

English . . .	0·51	Chemistry . . .	0·66
History . . .	0·73	Greek	0·78
French . . .	0·66	Mechanics . . .	0·61
Latin	0·67	Geography . .	0·52
Mathematics . .	0·72	Biology . . .	0·49
Physics . . .	0·63	German . . .	0·80

The above Table gives the averages obtained for four years over four Forms. A noticeable feature is that the values for English, Geography and Biology are low. This suggests that the examinations in

these subjects are not entirely satisfactory. On the other hand, the results in languages are relatively high, which seem to indicate that the technique of examining these subjects is more highly developed and more accurate. This also applies to Mathematics, but not to Physics, Chemistry and Mechanics. It seems probable that there is room for improvement in Science examinations. This may be due to lack of care in setting the papers or in marking these subjects.

The amount of ground to be covered is so great that the questions set can hardly test the whole ground and whole patches of work are left untested. Hence it is likely that the correlation between the examination result and the school record will not be too high. The moral seems to be that examiners should be most careful in spreading their questions over the syllabus, otherwise the result will not bear any true relation to the actual knowledge of the candidates. All question papers are based on the principle of sampling, and it is only when the sample is taken carefully that one can deduce the properties of the whole.

(2) *By Comparing the Standards.*

The standards of work to obtain a credit and a pass can be found from the published list of marks which is sent to the Schools. The exact percentage for a credit and the exact percentage for a pass is then known. These can be compared with the masters' estimates. It really amounts to this. Do the examiners and the masters draw the dividing lines between credit standard and pass standard and between pass standard and failure in the same place in a form ?

The Table on page 116 shows the comparison. In each case the number under C is the rank of the can-

didate obtaining a credit and the number under P the rank for a pass. Taking for example a rather good estimate in Form B Mathematics, we see that the master estimated that the 20th boy would get a credit and the 29th boy would get a pass. Actually, in the examination, the 24th boy got a credit and the 29th boy a pass. They were not necessarily the same boys as estimated, by the master—the correlation coefficient would show that—but it shows that it was a little easier to get a credit in the examination than the master estimated or alternatively, that the boys on the border-line for a credit did better than they were expected to do.

Such a Table kept over a space of years shows whether a master's estimates are reliable, and if not, where he is inclined to be too hard or too generous. It also acts as a safeguard in seeing that the examination produces reliable and satisfactory results, and whether the standard is being raised or lowered.

After some experience the estimates of masters are very reliable, and they should not be discarded without grave cause for the results of the examination. A master has the advantage over the examination that he has at least a year to judge the boys, whereas the examination may consist only of a paper of a few hours in which the questions may not adequately test the whole ground or work covered.

The above results can only be used to compare the estimate with the examination for a given Form.

The Forms cannot be compared unless the number of boys in each Form is taken into account and also allowance made for the numbers entered in each subject, as some boys in a Form may not offer all the subjects. Further, one Form may enter boys good at languages and another Form boys who are stronger at Science subjects.

COMPARISON OF RANKS FOR CREDIT OR PASS AND THE MASTER'S ESTIMATE

	English.		History.		French.		Latin.		Mathematics.		Physics.		Chemistry.	
	C.	P.	C.	P.	C.	P.	C.	P.	C.	P.	C.	P.	C.	P.
A. Exam. .	25th	26th	18½	24th	26th	—	26th	—	17th	25th	6th	11th	9th	17½
Estimate	15th	26th	8th	16th	26th	—	20th	—	23rd	26th	3rd	15th	8th	17th
B. Exam. .	26th	29th	18th	24th	21st	27th	15½	25th	24th	29th	26th	28th	18½	23½
Estimate	16th	30th	14th	27th	17th	30th	7th	17th	20th	29th	12th	25th	14th	27th
C. Exam. .	11th	27th	9th	19th	18th	25th	3rd	7th	23rd	26th	11th	17th	6th	11½
Estimate	10th	23rd	3rd	15th	9th	24th	5th	22nd	12th	23rd	4th	19th	4th	14th
D. Exam. .	9th	21½	9th	13th	4th	13th	1st	4th	11th	16th	6th	17th	2½	7th
Estimate	1st	13th	5th	12th	2nd	13th	1st	10th	1st	14th	—	11th	—	4th

APPENDIX

THE NORMAL CURVE

IF any trait or characteristic were measured for a large number of candidates, it would be found that the frequency curve obtained would be symmetrical. For instance, if the heights of all boys of a certain age were measured, then the frequency distribution would show that there were very few boys of a small height and a corresponding few with a big height, and as the height increases from the lowest height, the frequency of its occurrence would increase to a maximum, which would be found at the mean height. The frequency would then fall off until it became very small at the greatest height.

Such a normal curve is shown in Fig. 23.

This is a well-known curve in Mathematics, and is sometimes called the Probability Curve or the Gaussian Curve. It represents the distribution which would be found by chance or probability of any given trait amongst a large number of people. It has been found that the scores in a good Intelligence Test for a large number of candidates give a frequency curve which corresponds to the normal curve. This is because the frequency of any particular score will depend on the chance arrangement of a large number of factors, and for each score there will be some combination of these factors coming into play. For this reason, such a curve is called a probability curve, and the formula for such a curve can be deduced mathematically. Such a proof is beyond the scope of this book, but since the curve is of such importance, some of its main properties will be briefly stated.

(1) If n is the frequency of any score x on either side

of the central point, n_0 is the frequency of the central score or mean score, σ is a definite characteristic for the curve called the standard deviation, e is the exponential base of logarithms,

$$\text{Then} \qquad n = n_0 e^{-\frac{x^2}{2\sigma^2}}$$

(2) The characteristic σ defines the character of the curve.

If σ is big, then the curve is flat at the top.

If σ is small, then the curve is sharp and pointed.

Conversely, if the scores are very dispersed, the curve is well rounded and wide. The sides are not steep and σ is large. But if the scores are bunched together round the mean, then the curve is sharp and pointed and σ is small.

(3) At each end the curve approaches nearer and nearer to the X axis but never reaches it. As a matter of fact, however, it can be proved that if a distance of 3σ be measured on each side of the central point, only 0·2 per cent of the area of the curve will be outside. Since the area of the curve is proportional to the frequency, this means that all but 0·2 per cent of the scores will fall between 3σ below the centre point and 3σ above the centre point.

It can be further proved that by cutting off a distance of $2\frac{1}{2}\sigma$ on each side of the mean, about 1·24 per cent of the scores will be excluded. Therefore a base of length 6σ will embrace 99·8 per cent of the scores, and a base of length 5σ will embrace 98·76 per cent of the scores.

(4) Such a curve can be taken to give the ideal distribution of scores in an ideal test if the base be taken to represent the scores. If on the base a distance of 3σ is measured off on each side of the centre, and then this length of 6σ is divided into 100 parts, each of the points would correspond to a mark between 0 and 100. The height of the curve at each point would then give the number who obtained that particular score. This, of

course, is not strictly accurate, because 0·2 per cent of the area of the curve is excluded, but for all practical purposes it is quite accurate enough.

In the same way, if the base is cut off at $2\frac{1}{2}\sigma$ on each side of the mean, making a base length of 5σ, and then this base is divided into 100 parts, the ordinates at these points would give the number of candidates expected on the theory of probability to get each mark.

The result obtained by taking a base of 5σ is not so accurate as that obtained by taking a base of 6σ because 1·24 per cent of the cases are excluded in the former case compared with 0·2 per cent of the cases in the latter.

However, it is often more convenient to use a 5σ base than a 6σ base, because it lends itself to division into 10 or 100 parts more easily, and marks are so often given on a 10 point or a percentage scale. The further use of this curve and the frequency to be found at each per cent will be made in constructing a mark scale where due weight is given to the fact that some questions are harder than others.

THE IDEAL MARK SCALE

Imagine a perfectly normal probability curve to be drawn with its base divided into 5 equal parts, each equal to $\frac{1}{2}\sigma$, on each side of the centre line. The length of the base cut off will be 5σ and it will extend from $-2\frac{1}{2}\sigma$ to

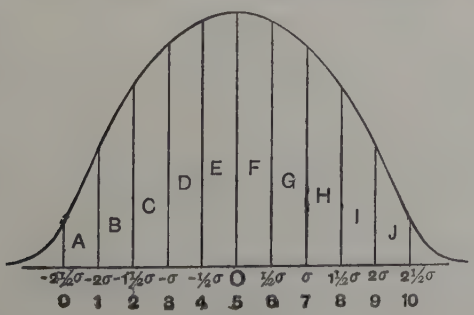

FIG. 23.

$+ 2\frac{1}{2}\sigma$. Let these points be marked 0, 1, 2, 3, etc., as in the figure, and let the ordinates be erected there.

It should be noted that by neglecting the extremities of the curve, we are neglecting 1·24 per cent. of the area of the curve. There is evidence to show that ability is distributed amongst mankind according to the probability curve, so the scale 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 may be taken to represent grades of ability and the ordinate at each point gives the number of persons of each grade of ability. Let us assume that an examination paper consisting of 10 properly graded problems has been drawn up so that on plotting the numbers solving each question we get a distribution curve identical with that in Fig. 23. This means that problem 1 is a measure of ability grade 1.

and „ „ 2 „ „ „ „ 2.

and so on.

Therefore the scale of difficulty of the problems is identical with the scale of ability.

Then area A represents the number who fail to solve question 1.

And area AB represents the number who fail to solve question 2.

„ „ ABC represents the number who fail to solve question 3.

„ „ ABCDEF represents the number who fail to solve question 6.

Now a person who solves all the problems up to and including 6 will have ability of the order 6. Hence if the marks are to be proportional to the difficulty of the problem and to the ability required to solve it, problem 6 will carry 6 marks.

Thus the mark value of any question will depend on the percentage of the candidates who fail to solve it, and this enables us to place any problem in its proper place on the mark scale.

From the properties of the probability curve it can be shown that the area ABCDEF is approximately 70 per cent. of the area of the whole curve. Hence problem 6

would be too hard for 70 per cent. of the candidates, but it would be solved by the remaining 30 per cent. who are represented by the area GHIJ.

Conversely if a problem is solved by 30 per cent. of the candidates that problem should rank as of difficulty 6 and carry 6 marks.

From Fig. 23 a percentile curve can be drawn which shows the percentages of candidates failing to solve each individual problem according to its grade of difficulty. Thus the percentage failing to solve question 6 of the 6th order of difficulty and carrying 6 marks is

$$\frac{\text{area ABCDEF}}{\text{whole area}} \times 100$$

and this will be represented by LM in the percentile curve (Fig. 24).

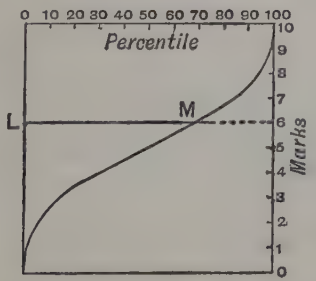

FIG. 24.

On the other hand, the dotted line M6 represents the percentage of candidates who solve question 6.

If a question is easy, only a few per cent will fail to solve it, so that this question should carry few marks. If a question is hard, many fail to solve it, so the question should carry many marks. Hence, if the percentage of the candidates failing to solve a given question is known, we can immediately, from the percentile curve, find the mark value of the question.

Fig. 24 is an example of an ideal percentile distribution calculated from the properties of the ideal frequency curve.

We see that if about 1·65 per cent of the candidates fail, the question should carry 1 mark. If approximately

6% of candidates fail in a question, that question should carry 2 marks.

15% „ „ „ „ „ „ question should carry 3 marks.

30% „ „ „ „ „ „ question should carry 4 marks.

50% „ „ „ „ „ „ question should carry 5 marks.

70% „ „ „ „ „ „ question should carry 6 marks.

85% „ „ „ „ „ „ question should carry 7 marks.

94% „ „ „ „ „ „ question should carry 8 marks.

98% „ „ „ „ „ „ question should carry 9 marks.

Hence we have the Table:

Percentage doing a question	2	6	15	30	50
Mark value of that question	9	8	7	6	5

Percentage doing a question	70	85	94	98·35
Mark value of that question	4	3	2	1

It is possible to construct a more refined table giving the figures to decimal points, but the above is sufficient to bring out the point.

The mark value of a question should vary with the hardness of the question. In this ideal scale, based on the probability curve, the fewer the number solving the question, the higher the marks, and conversely.

If this scale is to be applied, the procedure would be as follows. The questions for each candidate would be marked right or wrong. The total number of correct answers to each question would then be found and expressed as a percentage of the total number of candidates.

Suppose there were 60 candidates and 18 solved question No. 1. Then the percentage passing on this question is

$$\tfrac{1}{6}\tfrac{8}{0} \times 100 = 30 \text{ per cent.}$$

Hence the mark value of question No. 1 is $\underline{\underline{6}}$.

Other questions should be treated in the same way.

The marks a candidate has obtained are then totalled and the order of the candidate will be found. This order would almost certainly be different from the order produced by giving each question equal weight. Hence it is true to say that the mark weighting of the question will determine the final order. Thus it is of importance to weight each question properly. It may not be necessary to go to the above refinement, but it is necessary if a fair final order is to be obtained that some system of giving more marks for hard questions should be followed. The scale of the marks so adopted should bear some relation to the proportion of the candidates failing to do the question.

INDEX

Ability, 34, 35, 37, 120
Accuracy, 96, 98
Age Allowance, 104, 110
Age Group, 67
Analysis, 68, 73
Area, 24, 27, 50, 118
Association, 63
Average, 30, 79, 81

Bimodal, 31, 92
Borderline, 4, 111

Colligation, 63
Comparison, 66, 86, 114
Compounding, 11
Corrections, 43
Correlation, 8, 11, 39, 43, 110, 112
Credit, 8, 27, 114

Deviation, 29, 50
Difficulty, 120
Dispersion, 29, 53, 84, 92
Distribution, 3, 27, 73, 117

Estimates, 110, 114
Examinations, 1, 110

Footrule, 47
Form Marks, 13
Fourfold, 61
French, 99
Frequency, 24, 26, 31, 76, 117

Gaussian, 31, 117
General Paper, 13, 69, 103, 104
Grade, 28, 29
Graphs, 20, 23

Histogram, 24, 25, 73, 82
Homogeneous, 33, 95, 111

Ideal, 91, 92, 119, 121
Institute of Industrial Psychology, 35, 60
Integrated Frequency, 25, 27, 76
Intelligence Test, 6, 37, 82, 86
Interquartile, 29, 41, 72
I.Q., 34

Latin Paper, 88–95

Marking, 1, 9, 10, 11
Marks, 1, 4, 11, 19, etc.
Median, 28, 32, 79, 82, 86
Mode, 31

Norm, 33, 35, 36, 105
Normal Curve, 117

Objective, 37, 60
Order, 15, 21

Pass, 8, 114

Pearson, 61
P.E., 48, 50, 53, 55
Percentile, 25, 27, 32, 73, 80, 85, 91, 92, 102
Percentile Rank, 34, 102
Probability, 42, 117
Product Moment, 49, 55

Quartile, 29, 32, 51, 86
Quotient, 34

Ranks, 42, 55, 57, 102, 110
Record, 4, 37, 103
Regression, 41
Relationship, 61
Reliability, 58
Returns, 67

Sampling, 9, 114
Scaling, 17
Scholarships, 2, 12, 13, 104
School Certificate, 2, 3, 8, 27, 110
Science, 111, 114
Score Group, 73
Sets, 16
Sheppard's Method, 62
Skew, 26, 31
Slide Rule, 18, 19, 28
Spearman, 42, 47
Spread, 8
Standard Deviation, 41, 118, 119
Standard Tests, 6, 36
Standardization, 9
Subjective, 60

Tests, 36
Thermometer, 7
Time Table, 15

Unit Mark, 7
Unlike Signs, 62
Unreliability, 54
Unscaled, 22

Value, 121
Variability, 52

Weight, 8, 11, 12, 123

Made and Printed in Great Britain by Butler & Tanner Ltd., Frome and London

SCHOOL LABORATORY MANAGEMENT

By ARTHUR SUTCLIFFE, M.A., B.Sc., A.I.C.

Crown 8vo. 4s. 6d.

CONTENTS.

The Laboratory and its Fittings—Laboratory Organization and Management—The Preparation and Storage of Reagents—The Physical Laboratory—The Biological Laboratory—The Purchase of Apparatus and Material—Accidents in the Laboratory—Notes for the Laboratory Assistant in Chemistry—List of Bench Apparatus in Chemistry—Chart of Experiments in Practical Chemistry—Chart of Experiments in Practical Physics.

MR. FOX, IN HIS FOREWORD, SAYS :—

'The science master is faced with a set of practical difficulties which are peculiar to his subject. He has to look after apparatus, materials, chemicals, etc., and not only is he responsible for their upkeep but he has, in most cases, to purchase them or to advise about purchases. The detailed knowledge necessary for these purposes usually takes several years to acquire and the time and energy devoted is a serious tax on the science master's energy. Anything which tends to diminish it ought to be warmly welcomed.

For some years Mr. Sutcliffe has given a course on Laboratory Instruction to students at the Cambridge University Training College for Schoolmasters and this book deals with the substance of his course. All science masters will find information which is of the greatest practical moment to them. The advice given by Mr. Sutcliffe is pertinent, adequate and practical.'

¶ *As this is not a class text-book, specimen copies cannot be sent.*

English Language and Literature.

BY R. B. MORGAN, M.A., M.Litt.,

EDUCATION OFFICER, COUNTY BOROUGH OF CROYDON.

THE GROUNDWORK OF ENGLISH

An English Course on Modern Lines for Preparatory Schools and the Junior Forms of Secondary Schools. With Illustrations. In Two Parts. Second Impression. *1s. 6d. each.*

Part I consists mainly of Oral Exercises, Part II of Written Work.

A JUNIOR ENGLISH GRAMMAR

162nd Thousand. Cloth boards, *2s.*

This grammar is now too well known to need any detailed description. Though it has passed through many editions, and a few minor alterations have been made in it from time to time in response to suggestions from teachers who have used it, it is practically the same as on the day it was published ; this shows on what sound lines it was conceived in the first instance.

EXERCISES IN ENGLISH GRAMMAR

Third Impression. *1s. 9d.*

A companion to the above, which can be used either independently or as a supplement to it.

A SENIOR ENGLISH GRAMMAR

By R. B. Morgan, M.A., M.Litt., and H. A. Treble, M.A., Senior English Master, Selhurst Grammar School. Third Impression. *3s. 6d.*

This book is intended to follow and supplement Morgan's "Junior English Grammar," and to provide a trustworthy manual for all examinations which include a grammar test. One of its principal features is the large number of quotations from standard authors.

THE TEACHING OF GRAMMAR

A Book for Teachers and Students, dealing with the scientific side of the subject. By Laura Brackenbury, M.A., formerly Principal of the Clapham Training College. *2s. 6d. net.*

A HIGHER COURSE OF ENGLISH PRACTICE

A Guide to Literary Appreciation and Expression. By R. B. Morgan, M.A., M.Litt., and R. B. Lattimer, M.A. Third Impression. ***3s. 6d.***

This book has been written for those who have reached the standard of the final School Leaving Examination and University Matriculation.

Every chapter has been based on a long experience of pupils' difficulties and frailties, and almost every question has done duty in some examination or other.

For Central Schools.

ENGLISH WORDS AND THEIR USES

By R. B. Morgan, M.A., M.Litt., and R. B. Lattimer, M.A. ***1s.***

A reprint of the first four chapters of the above book, for the use of younger pupils.

THE ELEMENTS OF GRAMMAR *1s.*

An edition of Part I of Morgan's "Junior English Grammar." Second Impression.

EXERCISES IN THE ELEMENTS OF GRAMMAR *1s.*

A companion book to the above, consisting of Part I of "Exercises in English Grammar."

BY E. E. KITCHENER, M.A.,

WHITGIFT GRAMMAR SCHOOL, CROYDON.

ENGLISH COMPOSITION

37th Thousand. *2s.*

Especially suitable for School Certificate and similar Examinations.

This book is a first book in English Composition suitable for pupils between the ages of eight and thirteen.

EXERCISES IN ENGLISH COMPOSITION

Third Impression. *1s. 9d.*

For teachers who feel the need of a collection of additional material for practice.

The English Tradition of Education
by CYRIL NORWOOD, M.A., D.Litt., Headmaster of Harrow School. *2nd Impression* 10s. 6d. net.

Books by Ernest Weekley, M.A., Nottingham University.

The Romance of Words. *7th Impression* 6s. net.

Words Ancient and Modern. *2nd Impression* 5s. net.

More Words Ancient and Modern 5s. net.

The Romance of Names. *3rd Edition* 6s. net.

Surnames. *3rd Edition* 6s. net.

Etymological Dictionary £2 2s. net.

A Concise Etymological Dictionary of Modern English 7s. 6d. net.

Adjectives and Other Words 5s. net.

Books by Henry Cecil Wyld, Merton Professor, Oxford.

A Short History of English. *3rd (enlarged) Edition* 10s. 6d. net.

The Historical Study of the Mother Tongue. *4th Impression.* 9s. net.

The Growth of English. *6th Impression* 5s. net.

The Teaching of Reading in Training Colleges. *3rd Impression.* 3s. net.

Studies in English Rhymes 5s. net.

The Place of the Mother Tongue in National Education 1s. net.

A First Course in Précis Writing
by T. W. MOLES, B.A., B.Sc., Rutherford College 2s.

Short Studies in Shakespeare
by G. F. BRADBY 6s. net.

JOHN MURRAY, ALBEMARLE STREET, LONDON, W.1